

Integral and Functional Analysis

(Updated Edition)

$\{a, b, c\}$

$i \equiv \ddot{O} - 1$

\mathbb{R} = the set of real numbers

$x \hat{X}$

Jie Xiao

MATHEMATICS RESEARCH DEVELOPMENTS

NOVA

MATHEMATICS RESEARCH DEVELOPMENTS

**INTEGRAL AND FUNCTIONAL
ANALYSIS
(UPDATED EDITION)**

No part of this digital document may be reproduced, stored in a retrieval system or transmitted in any form or by any means. The publisher has taken reasonable care in the preparation of this digital document, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained herein. This digital document is sold with the clear understanding that the publisher is not engaged in rendering legal, medical or any other professional services.

MATHEMATICS RESEARCH DEVELOPMENTS

Additional books and e-books in this series can be found
on Nova's website under the Series tab.

MATHEMATICS RESEARCH DEVELOPMENTS

**INTEGRAL AND FUNCTIONAL
ANALYSIS
(UPDATED EDITION)**

JIE XIAO



Copyright © 2021 by Nova Science Publishers, Inc.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means: electronic, electrostatic, magnetic, tape, mechanical photocopying, recording or otherwise without the written permission of the Publisher.

We have partnered with Copyright Clearance Center to make it easy for you to obtain permissions to reuse content from this publication. Simply navigate to this publication's page on Nova's website and locate the "Get Permission" button below the title description. This button is linked directly to the title's permission page on copyright.com. Alternatively, you can visit copyright.com and search by title, ISBN, or ISSN.

For further questions about using the service on copyright.com, please contact:

Copyright Clearance Center

Phone: +1-(978) 750-8400

Fax: +1-(978) 750-4470

E-mail: info@copyright.com.

NOTICE TO THE READER

The Publisher has taken reasonable care in the preparation of this book, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained in this book. The Publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance upon, this material. Any parts of this book based on government reports are so indicated and copyright is claimed for those parts to the extent applicable to compilations of such works.

Independent verification should be sought for any data, advice or recommendations contained in this book. In addition, no responsibility is assumed by the Publisher for any injury and/or damage to persons or property arising from any methods, products, instructions, ideas or otherwise contained in this publication.

This publication is designed to provide accurate and authoritative information with regard to the subject matter covered herein. It is sold with the clear understanding that the Publisher is not engaged in rendering legal or any other professional services. If legal or any other expert assistance is required, the services of a competent person should be sought. FROM A DECLARATION OF PARTICIPANTS JOINTLY ADOPTED BY A COMMITTEE OF THE AMERICAN BAR ASSOCIATION AND A COMMITTEE OF PUBLISHERS.

Additional color graphics may be available in the e-book version of this book.

Library of Congress Cataloging-in-Publication Data

ISBN: ; 9: /3/7583; /839/9*gDqqm-

Published by Nova Science Publishers, Inc. † New York

Contents

Preface	ix
Acknowledgments	xi
1 Preliminaries	1
1.1 Sets, Relations, Functions, Cardinals and Ordinals	1
1.2 Reals, Some Basic Theorems and Sequence Limits	10
Problems	19
2 Riemann Integrals	21
2.1 Definitions, Examples and Basic Properties	21
2.2 Algebraic Operations and the Darboux Criterion	28
2.3 Fundamental Theorem of Calculus	35
2.4 Improper Integrals	43
Problems	52
3 Riemann-Stieltjes Integrals	57
3.1 Functions of Bounded Variation	57
3.2 Definition and Basic Properties	61
3.3 Nonexistence and Existence for Integrals	65
3.4 Evaluations of Integrals	74
3.5 Improper Situations	78
Problems	83
4 Lebesgue-Radon-Stieltjes Integrals	87
4.1 Foundational Material	88
4.2 Essential Properties	95

4.3	Convergence Theorems	100
4.4	Extension via Measurability	107
4.5	Double, Iterated and Generic Integrals	113
	Problems	128
5	Absolute Continuities	
	in Lebesgue Integrals	135
5.1	Lebesgue's Outer Measure and Vitali's Covering	135
5.2	Derivatives of Increasing Functions	143
5.3	Absolutely Continuous Functions	148
5.4	Cantor's Ternary Set and Singular Function	158
5.5	Lebesgue's Points	163
	Problems	169
6	Metric Spaces	173
6.1	Metrizable Topology and Connectedness	173
6.2	Completeness	181
6.3	Compactness, Density and Separability	187
	Problems	195
7	Continuous Mappings	199
7.1	Criteria for Continuity	199
7.2	Continuous Mappings over Compact or Connected Metric Spaces	204
7.3	Sequences of Mappings	207
7.4	Contractions	212
7.5	Structures of Metric Spaces	218
	Problems	221
8	Normed Linear Spaces	225
8.1	Linear Spaces, Norms and Quotient Spaces	225
8.2	Finite Dimensional Spaces	233
8.3	Bounded Linear Operators	239
8.4	Linear Functionals via Hahn-Banach Extension	243
	Problems	249

9	Banach Spaces via Operators and Functionals	255
9.1	Definition and Beginning Examples	255
9.2	Uniform Boundedness - Open Map - Closed Graph	261
9.3	Dual Banach Spaces by Examples	267
9.4	Weak and Weak* Topologies	281
9.5	Compact and Dual Operators	288
	Problems	292
10	Hilbert Spaces and Their Operators	297
10.1	Definition, Examples and Basic Properties	297
10.2	Orthogonality, Orthogonal Complement and Duality	301
10.3	Orthonormal Sets and Bases	304
10.4	Five Special Bounded Operators	311
10.5	Compact Operators via Spectrum	321
	Problems	332
	Hints or Solutions	337
	References	389
	About the Author	393
	Index	395

Preface

Since the publication of the textbook, *Integral and Functional Analysis (IaFA)*, in 2008, many helpful comments from colleagues and students have been received. Not only were typos corrected, but also some interesting topics were added for enhancements. Below is a revision of the original preface.

This book is based on three closely-related courses: (a) Lebesgue Integration; (b) Integration and Metric Spaces; (c) Functional Analysis, which I have offered at Memorial University since 2002. Though the part on Functional Analysis has been used for both an undergraduate course and a graduate course, this textbook is designed primarily for senior undergraduate students. The prerequisites of this textbook are deliberately modest, and it is assumed that the students have some familiarity with Introductory Calculus and Linear Algebra plus the basic (direct, indirect) proof methods.

I have striven to give an exposition which is at the same time introductory and modern in spirit, yet which addresses itself to the classical concerns of these three Analysis courses in mathematics. This approach leads to my aim in writing this textbook: I must present the basic ideas and results of this area in a natural sequence. On the other hand, enough material is covered to provide a firm base on which to build for later studies in Advanced Analysis and Partial Differential Equations.

The textbook comprises two parts: (i) Integral Analysis; (ii) Functional Analysis, with results from the first part being used to partially motivate problems discussed in the second part, which can be covered from beginning to end in two semesters.

The first part (Chapters 1, 2, 3, 4, 5) – Integral Analysis – a mathematical theory that defines the reasonable integrals for functions of different kinds. The integral of an individual function is a generalization of area, mass, volume, sum, and total. There are several possible definitions of integration with differ-

ent technical underpinnings. In this part, we start with the basic arithmetic and topological properties of the real number system. Then we discuss Riemann integration including some essential properties and the fundamental theorem of calculus, and use it as a model to handle both Riemann-Stieltjes integration via bounded variation and Lebesgue-Radon-Stieltjes integration covering three convergence theorems and Fubini-Tonelli's theorem as well as absolutely continuous functions within Lebesgue integration.

The second part (Chapters 6, 7, 8, 9, 10) – Functional Analysis – a mathematical theory that is concerned with function spaces and their functionals and operators. It has its historical roots in the investigation of transformations such as the Fourier transform and in the study of differential and integral equations. The functional goes back to the calculus of variations, implying a function whose argument is a function. In this part, motivated by the first part, we first work with more abstract metric spaces and their continuous mappings, and next introduce elements of the functional theory including Banach spaces, and Hilbert spaces whose norms arise from inner products. Of course, we also discuss major and foundational results of this theory such as the Hahn-Banach theorem extending a functional from a subspace to the full space in a norm-preserving fashion, the uniform boundedness principle, the open mapping theorem and closed graph theorem, and the approximation for compact operators on the complex Hilbert spaces.

In the process of learning IaFA, the students are strongly suggested to follow such an important principle that the best way to learn mathematics is to do mathematics. Moreover, the students are urged to acquire the habit of studying with paper and pen/pencil in hand (plus a desktop/laptop computer); in this way mathematics will become increasingly meaningful to them. Each chapter is followed by a series of exercises dealing with the material presented in that chapter. Although brief hints or complete solutions for these exercises are included, the students are recommended to at least do most of the exercises before checking their solutions or hints. A list of references is provided – those textbooks also supplement this textbook since they were used while preparing the courses. Last but not least, an index is attached.

Jie Xiao

September 2020 to January 2021

Acknowledgments

I wish to thank all the colleagues whoever supported the preparation of the original textbook and any subsequent revisions. Next, I am grateful to the students of the Department of Mathematics and Statistics at Memorial University, who actively participated in the three courses and gave some helpful comments and suggestions on the textbook. Last but not least, I would like to thank Nova Science Publishers for continuing to publish this updated textbook.

Chapter 1

Preliminaries

The objective of this chapter of two sections is to set up an appropriate groundwork for developing the theory of integrals and functionals. The first section establishes some conventions concerning sets, relations, functions, and cardinal and ordinal numbers. The second section is devoted to collecting the basic properties (as axioms) of the real numbers including their arithmetic, order and least upper bound, as well as limit rules.

1.1 Sets, Relations, Functions, Cardinals and Ordinals

A set is a collection or group of objects, considered as an entity unto itself. Sets are usually symbolized by uppercase, italicized, boldface letters such as A , B , X , Y , or Z . Each object in a set is called a member or an element of the set. If x is an element of the set X , then we write $x \in X$, and hence we denote by $x \notin X$ when x is not an element of X . Sets may be expressed either by explicitly listing their elements inside of braces – for example, $\{a, b, c\}$ is the set having a , b , and c as members, or by giving the property that every element of the set possesses and which is possessed by no member not in the set – for instance, $\{x : x \text{ is a car}\}$ is the set of all cars, or the set of all x such that x is a car. Below are the standard sets that will be used throughout this textbook:

- \mathbf{N} = the set of natural numbers = $\{1, 2, 3, \dots, n, \dots\}$;
 \mathbf{Z} = the set of integers = $\{0, 1, -1, 2, -2, 3, -3, \dots, n, -n, \dots\}$;
 \mathbf{Q} = the set of rational numbers = $\{mn^{-1} : m \text{ and } n \text{ are integers with } n \neq 0\}$;
 \mathbf{R} = the set of real numbers;
 \mathbf{C} = the set of complex numbers.

Remark 1.1.1.

- (i) In \mathbf{Q} we identify mn^{-1} and kl^{-1} if $lm = kn$.
 (ii) $\mathbf{C} = \mathbf{R} + i\mathbf{R} = \{z = x + iy : x, y \in \mathbf{R}\}$, where $i = \sqrt{-1}$ is the imaginary unit; $x = \Re z$ and $y = \Im z$ are called the real and imaginary parts of $z = x + iy$ respectively.

Definition 1.1.2. Given two sets X and Y . We say that:

- (i) X is a subset of Y , written $X \subseteq Y$, provided that $x \in X$ implies $x \in Y$ – clearly $X \not\subseteq Y$ denotes that X is not a subset of Y ;
 (ii) X and Y are equal, denoted $X = Y$, provided that $X \subseteq Y$ and $Y \subseteq X$ – clearly, $X \neq Y$ means that X is not equal to Y ;
 (iii) X is a proper subset of Y , written $X \subset Y$, provided that $X \subseteq Y$ but $X \neq Y$;
 (iv) $X \cup Y$, $X \cap Y$, and $X \setminus Y$ are the union of X and Y , i.e., the set $\{x : x \in X \text{ or } x \in Y\}$, the intersection of X and Y , i.e., the set $\{x : x \in X \text{ and } x \in Y\}$, and the complement of Y in X , i.e., the set $\{x : x \in X \text{ and } x \notin Y\}$.
 (v) \emptyset is the empty set, i.e., the set which contains no elements – moreover, X and Y are disjoint whenever $X \cap Y = \emptyset$.

As the two most basic set operations, union and intersection can be used for any family of sets. More precisely, if $\{X_j\}_{j \in I}$ is any family of sets indexed by some set I which may be considered simply as a set of labels for distinguishing the various members of the family of sets, then

$$\cup_{j \in I} X_j = \{x : x \in X_j \text{ for some } j \in I\} \quad \text{and} \quad \cap_{j \in I} X_j = \{x : x \in X_j \text{ for all } j \in I\}$$

stand for the union and intersection of $\{X_j\}_{j \in I}$ respectively. Acting with the complement operation, the union and intersection satisfy the DeMorgan law for an arbitrary family of sets as follows.

Proposition 1.1.3. *Given a set X , let $\{X_j\}_{j \in I}$ be a family of sets indexed by I . Then*

$$\bigcup_{j \in I} (X \setminus X_j) = X \setminus \bigcap_{j \in I} X_j \quad \text{and} \quad \bigcap_{j \in I} (X \setminus X_j) = X \setminus \bigcup_{j \in I} X_j.$$

Proof. It suffices to verify the left-hand equality since the right-hand equality can be proved similarly. First, suppose that $x \in \bigcup_{j \in I} (X \setminus X_j)$. Then $x \in X \setminus X_j$ for some $j \in I$ and hence $x \notin X_j$, implying $x \notin \bigcap_{j \in I} X_j$. Consequently, $x \in X \setminus \bigcap_{j \in I} X_j$, namely,

$$\bigcup_{j \in I} (X \setminus X_j) \subseteq X \setminus \bigcap_{j \in I} X_j.$$

On the other hand, if $x \in X \setminus \bigcap_{j \in I} X_j$ then $x \in X$ but $x \notin \bigcap_{j \in I} X_j$. Accordingly, there exists some $j \in I$ such that $x \in X \setminus X_j$. Then $x \in \bigcup_{j \in I} X \setminus X_j$. Consequently,

$$X \setminus \bigcap_{j \in I} X_j \subseteq \bigcup_{j \in I} (X \setminus X_j).$$

According to Definition 1.1.2 (ii), we derive the desired equality. \square

Next, we introduce the Cartesian product and relation, and equivalence relation.

Definition 1.1.4. *Given two sets X and Y .*

(i) *The Cartesian product $X \times Y$ of X and Y is defined to be the set of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$ – here the ordered pair (x, y) is the set whose elements are $\{x\}, \{x, y\}$, in symbols $(x, y) = \{\{x\}, \{x, y\}\}$.*

(ii) *A subset R of $X \times Y$ is called a relation between X and Y – in particular a subset of $X \times X$ is said to be a relation on X – in this case, $(x, y) \in R$ may also be denoted by xRy or x being R -related to y .*

(iii) *A relation R on X is an equivalence relation provided that it has the following properties for all $x, y, z \in X$:*

- (a) xRx (reflexivity);
- (b) xRy implies yRx (symmetry);
- (c) xRy and yRz yield xRz (transitivity).

Here, for simplicity we have chosen (x, y) as the notation of an ordered pair, but hope that this symbolic choice will not make any trouble whenever the same notation is used later on to stand for an open interval of reals.

Of course, the equality “=” on any set X is an equivalence relation. To better understand this definition, we prove the following proposition.

Proposition 1.1.5. *Let $(x, y), (x', y') \in X \times Y$. Then $(x, y) = (x', y')$ if and only if $x = x'$ and $y = y'$.*

Proof. Clearly, $x = x'$ and $y = y'$ imply

$$(x, y) = \{\{x\}, \{x, y\}\} = \{\{x'\}, \{x', y'\}\} = (x', y').$$

Conversely, suppose that $(x, y) = (x', y')$. To prove $x = x'$ and $y = y'$, we consider two cases.

Case 1: $x = y$. Then $\{x\} = \{x, y\}$ and hence $(x', y') = (x, y) = \{\{x\}\}$. This means that $\{\{x'\}, \{x', y'\}\}$ has only one element $\{\{x\}\}$ and so that $\{x', y'\} = \{x'\} = \{x\} = \{y\}$. Accordingly, $x' = y'$.

Case 2: $x \neq y$. Then the argument for *Case 1* yields $x' \neq y'$. Since $(x, y) = (x', y')$, we conclude $\{x\} \in \{\{x'\}, \{x', y'\}\}$ and so either $\{x\} = \{x'\}$ or $\{x\} = \{x', y'\}$. This in turns implies $x' \in \{x\}$ and then $x = x'$. Similarly, $\{x, y\} \in \{\{x'\}, \{x', y'\}\}$ yields either $\{x, y\} = \{x'\}$ or $\{x, y\} = \{x', y'\}$. But since $x \neq y$, we must have $\{x, y\} = \{x', y'\}$. Now $x = x'$, $x \neq y$ and $y \in \{x', y'\}$, it follows that $y = y'$, as desired. \square

As a special relation, we have the notion of a function.

Definition 1.1.6. *Given two sets X and Y .*

(i) *A function f between X and Y is a nonempty relation $f \subseteq X \times Y$ such that if $(x, y), (x, y') \in f$ then $y = y'$. The following sets*

$$D(f) = \{x \in X : \text{there is a } y \in Y \text{ such that } (x, y) \in f\}$$

and

$$R(f) = \{y \in Y : \text{there is an } x \in X \text{ such that } (x, y) \in f\}$$

are called the domain and the range of f respectively. In the case of $D(f) = X$, f is called a function from X into Y , denoted $f : X \rightarrow Y$, with

$$f(A) = \{f(x) : x \in A\} \quad \text{and} \quad f^{-1}(B) = \{x \in X : y = f(x) \in B\}$$

as the image of $A \subseteq X$ and the pre-image of $B \subseteq Y$ respectively.

(ii) *A function $f : X \rightarrow Y$ is called onto or surjective provided $f(X) = Y$.*

(iii) *A function $f : X \rightarrow Y$ is called one-to-one or injective provided that $f(x) = f(x')$ for any $x, x' \in X$ implies $x = x'$.*

(iv) A surjection is an onto function, while an injection is a one-to-one function. Furthermore, a function which is both onto and one-to-one is said to be bijective or a bijection.

Proposition 1.1.7. *Given two sets X and Y , let $f : X \rightarrow Y$ be a function.*

(i) *Suppose that $A, B \subseteq X$. Then $f(A \cup B) = f(A) \cup f(B)$. Furthermore $f(A \cap B) = f(A) \cap f(B)$ whenever f is injective.*

(ii) *Suppose that $A \subseteq X$ and $B \subseteq Y$. Then $f^{-1}(f(A)) = A$ respectively $f(f^{-1}(B)) = B$ provided f is injective respectively surjective.*

Proof. (i) If $y \in f(A \cup B)$, then by definition there is an $x \in A \cup B$ such that $y = f(x)$. Now $x \in A$ or $x \in B$, consequently $f(x) \in f(A)$ or $f(x) \in f(B)$. In either case, $y \in f(A) \cup f(B)$. This proves $f(A \cup B) \subseteq f(A) \cup f(B)$. On the other hand, if $y \in f(A) \cup f(B)$, then $y \in f(A)$ or $y \in f(B)$ and hence there exists $x \in A \cup B$ such that $y = f(x)$, namely, $y \in f(A \cup B)$. Thus, $f(A) \cup f(B) \subseteq f(A \cup B)$. Accordingly, $f(A \cup B) = f(A) \cup f(B)$.

Clearly, $f(A \cap B) \subseteq f(A) \cap f(B)$. Now if f is injective, then for any $y \in f(A) \cap f(B)$ there exist $x \in A$ and $x' \in B$ such that $y = f(x) = f(x')$, and hence $x = x' \in A \cap B$. Accordingly, $y \in f(A \cap B)$, i.e., $f(A) \cap f(B) \subseteq f(A \cap B)$, whence giving $f(A \cap B) = f(A) \cap f(B)$.

(ii) This follows immediately from Definition 1.1.6 (i). □

Remark 1.1.8. *We make two comments on Proposition 1.1.7 (ii).*

(i) *The symbol $f^{-1}(B)$ is not to be thought of as an inverse function applied to points in B . But, if $f^{-1}(y) = \{x \in X : f(x) = y\}$ (for each $y \in Y$) contains precisely one member in X , then f^{-1} defines a function from Y into X . In fact, this amounts to $f : X \rightarrow Y$ being bijective. To see this, if f is a bijection, then for any $y \in Y$ there exists one (surjective) and only one (injective) element $x \in X$ such that $y = f(x)$, and hence $f^{-1}(y) = x$ is uniquely defined for each $y \in Y$. Consequently, f^{-1} is a function from Y into X . Conversely, if $f^{-1} : Y \rightarrow X$ is a function, then for each $y \in Y$, $f^{-1}(y) = x$ is a unique member of X , that is to say, f is injective. Moreover, $f^{-1}(y) = x$ produces $f(x) = y$ and so that f is surjective. Therefore, f is bijective.*

(ii) *The symbols $f^{-1}(f(A))$ and $f(f^{-1}(B))$ induce the notion of composition of functions. For functions $f : X \rightarrow Y$ and $g : Y \rightarrow Z$, we say that $g \circ f$, the composition of g with f , is a function from X into Z determined by $(g \circ f)(x) = g(f(x))$ for each $x \in X$. In order to make this meaningful, we naturally require*

$f(X) \subseteq Y$. It is evident to see that if $f : X \rightarrow Y$ is bijective then $f^{-1} : Y \rightarrow X$ is bijective and $f^{-1} \circ f = i_X$ and $f \circ f^{-1} = i_Y$ where $i_X : X \rightarrow X$ and $i_Y : Y \rightarrow Y$ stand for the identity functions:

$$i_X(x) = x \text{ for } x \in X \text{ and } i_Y(y) = y \text{ for } y \in Y.$$

We now introduce the cardinality of a set, which is a property describing the size of the set by describing it using a cardinal number. This is actually equivalent to comparing two sets, and so leads to the following concept.

Definition 1.1.9. Two sets X and Y are called equivalent, or said to have the same cardinality, denoted $X \sim_R Y$, provided that there exists a bijection between X and Y . In particular, any set that has the same cardinality \aleph_0 as the set of all natural numbers \mathbf{N} is said to be an infinite countable set, and if a set is \emptyset or there is a bijection between this set and $\{1, 2, \dots, n\}$ for some $n \in \mathbf{N}$ of then it is said to be finite; otherwise, the set is said to be uncountable.

Example 1.1.10.

- (i) If $X = \{a, b, c\}$ and $Y = \{\text{tables, chairs, chalkboards}\}$, then they both have three elements.
- (ii) The set $2\mathbf{N} = \{2, 4, 6, \dots, 2n, \dots\}$ has the same cardinality as \mathbf{N} since the function $f(n) = 2n$, is a bijection between \mathbf{N} and $2\mathbf{N}$.
- (iii) \mathbf{Z} has the cardinality \aleph_0 . In fact, if

$$\mathbf{Z}_- = \{-1, -2, -3, \dots, -n, \dots\}$$

and

$$f(x) = \begin{cases} 2x & , x \in \mathbf{N} \\ 1 & , x = 0 \\ (-1)(2x) + 1 & , x \in \mathbf{Z}_- \end{cases},$$

then $\mathbf{Z} = \mathbf{N} \cup \{0\} \cup \mathbf{Z}_-$ and f is a bijection between \mathbf{N} and \mathbf{Z} .

Proposition 1.1.11.

- (i) Any subset of a finite set is finite.
- (ii) Any subset of any countable set is countable.
- (iii) The union of any countable collection of countable sets is countable.
- (iv) The Cartesian product of any two countable sets is countable.

Proof. (i) Suppose X is finite and $A \subseteq X$. Then either X is \emptyset or there exists a bijection $f : X \rightarrow \{1, 2, 3, \dots, n\}$ for some $n \in \mathbf{N}$. If $A = \emptyset$, then it is finite. If $A \neq \emptyset$, then the assumption $A \subseteq X$ yields

$$f(A) \subseteq \{1, 2, 3, \dots, n\},$$

and hence we may assume

$$f(A) = \{j_1, j_2, j_3, \dots, j_m\}.$$

Consequently, upon defining

$$g : A \rightarrow \{1, 2, 3, \dots, m\} \text{ by } g(x) = k \text{ where } f(x) = j_k \text{ for } x \in A,$$

we see that A is finite.

(ii) Assume that X is countable and $S \subseteq X$. Then $X = \{x_j\}_{j=1}^{\infty}$. If $S = \emptyset$, then it is countable. If $S \neq \emptyset$, then it contains at least one element, say s . For each $j \in \mathbf{N}$, set $y_j = x_j$ or $y_j = s$ when $x_j \in S$ or $x_j \notin S$. Then this establishes a bijection between S and $\{y_j\}_{j=1}^{\infty}$. Thus S is countable.

(iii) Suppose that $\{A_j\}_{j=1}^{\infty}$ is a countable collection of countable sets $A_j = \{a_{jk}\}_{k=1}^{\infty}$. Define $f(a_{jk}) = 2^j 3^k$. Then f is a bijection between $\cup_{j=1}^{\infty} A_j$ and a subset of \mathbf{N} . Thus $\cup_{j=1}^{\infty} A_j$ is countable due to (ii).

(iv) Assume that $X = \{x_j\}_{j=1}^{\infty}$ and $Y = \{y_j\}_{j=1}^{\infty}$. Then each member of $X \times Y$ has the form (x_j, y_k) for $j, k \in \mathbf{N}$. Define $f : X \times Y \rightarrow \mathbf{N}$ by $f(x_j, y_k) = 2^j 3^k$. Then f is one-to-one. Thus $X \times Y$ is countable. \square

Example 1.1.12.

(i) **Q** is countable. In fact, via associating mn^{-1} , $n \neq 0$, m and n relatively prime, we see that **Q** has the same cardinality as a subset of $\mathbf{Z} \times \mathbf{Z}$ which is a countable set by (iv).

(ii) **R** is uncountable. To verify this, let X be the set of unending decimals between 0 and 1 and that contain only 0 and 1 as digits. If f is an arbitrary function from \mathbf{N} into X , then f cannot be onto, and hence X is uncountable. In fact, we form a member $d = 0.x_1x_2x_3x_4 \dots$ of X as follows: If the first digit of $f(1)$ is 0 or 1, then x_1 is taken to be 1 or 0, and consequently, d differs from $f(1)$ in the first digit. We choose x_2 to be either 0 or 1, but different from the second digit of $f(2)$. Via continuing this process, we find that d is not $f(n)$ for any n since d differs from $f(n)$ for any n at least in the n th digit. Consequently,

$d \notin f(\mathbf{N})$, and so f is not surjective. Note that X is a subset of \mathbf{R} . So, if \mathbf{R} were countable, then X would also be countable by Proposition 1.1.11 (ii). This contradiction proves that \mathbf{R} is uncountable.

The preceding discussion on the cardinals of sets reveals

$$X = \{1, 2, 3, \dots\} \sim_R Y = \{\dots, 3, 2, 1\}.$$

However, if X is ordered by the usual relation \leq and Y ordered by \geq then X has a first but not a last element, and Y has a last but not a first, and hence this gives a way to distinguish both sets. So, we are finally led to a brief consideration of the ordinals of sets.

Definition 1.1.13. Let \preceq be a relation on the given set X .

(i) \preceq is said to be a partial order on X provided the following three properties hold for $x, y, z \in X$:

- (a) $x \preceq x$ (reflexivity);
- (b) $x \preceq y$ and $y \preceq x$ imply $x = y$ (antisymmetry);
- (c) $x \preceq y$ and $y \preceq z$ yield $x \preceq z$ (transitivity).

In this case, (X, \preceq) stands for a partially ordered set in which not all pairs of elements are mutually comparable.

(ii) \preceq is said to be a total/linear order provided the following three properties hold for $x, y, z \in X$:

- (a) $x \preceq y$ or $y \preceq x$ (totalness);
- (b) $x \preceq y$ and $y \preceq x$ imply $x = y$ (antisymmetry);
- (c) $x \preceq y$ and $y \preceq z$ yield $x \preceq z$ (transitivity).

In this case, (X, \preceq) stands for a totally/linearly ordered set, or a chain. The totalness property can be stated thus that any pair of elements in the chain are mutually comparable.

Notice that the totalness condition implies reflexivity. Thus a total order is also a partial order, that is, a binary relation which is reflexive, antisymmetric and transitive. It follows that a total order can also be defined as a partial order that is total.

Example 1.1.14.

(i) Given a set X , let $\mathcal{P}(X)$ be the collection of all subsets of X . Then $(\mathcal{P}(X), \subseteq)$ is a partially ordered set, but usually it is not totally ordered.

(ii) (\mathbf{Z}, \leq) is totally ordered.

Definition 1.1.15. Let (X, \preceq) be a partially ordered set.

(i) An element u of X is an upper bound of a subset A of X , provided that $a \preceq u$ for all elements $a \in A$ – in this case, A is called bounded from above. Using $u \preceq a$ instead of $a \preceq u$ leads to the definition of a lower bound of A , i.e., A being bounded from below.

(ii) An element m of a subset A of X is a maximal element of A if $m \preceq a$ for some $a \in A$ implies $m = a$. The definition for a minimal element is obtained by using $a \preceq m$ instead of $m \preceq a$.

Clearly, a subset of a partially ordered set may fail to have any upper bounds. Consider for example the subsets of \mathbf{N} which are greater than a given natural number. On the other hand, a set may have many several upper and lower bounds, and hence we are usually interested in picking out specific elements from the sets of upper or lower bounds. This leads to the consideration of least upper bounds (or suprema) and greatest lower bounds (or infima).

What is important to note about maximal elements is that they are in general not the greatest elements of a subset S . Indeed, consider $\mathcal{P}(\mathbf{N})$ – the set of all subsets of the natural numbers ordered by subset inclusion. The subset S of all one-element sets of \mathbf{N} consists only of maximal elements, but has no greatest element. This example also shows that maximal elements are usually not unique and that it is well possible for an element to be both maximal and minimal at the same time.

If a subset has a greatest element, then this is the unique maximal element. But, on the other hand, even if a set has only one maximal element, it is not necessarily the greatest one. Take the set of natural numbers in their usual order, which obviously has no maximal elements, and add a single new element a which can only be compared to itself, in other words, it is neither smaller nor greater than any natural number. Then the whole set has a as a single maximal element that is not the greatest element. Similar conclusions are valid for minimal elements. Nevertheless, in a totally ordered set, the terms maximal (minimal) element and greatest (least) element are the same, which is why both terms are used interchangeably in fields like analysis where only total orders are considered.

With the previous concepts, we may state the celebrated Zorn's lemma, also known as the Kazimierz Kuratowski - Max Zorn lemma, as follows.

Zorn's Lemma. *Every partially ordered set in which every totally ordered subset has an upper bound contains at least one maximal element.*

We offer no proof for this lemma because it is equivalent to the following axiom of choice in the sense that either one is sufficient to prove the other:

The Axiom of Choice. *Given an arbitrary collection $\{A_j\}_{j \in I} = \{A_j : j \in I\}$ of nonempty sets indexed by the nonempty set I , there is a function $f : I \rightarrow \bigcup_{j \in I} A_j$, called a choice or selection function, such that $f(j) \in A_j$ when $j \in I$.*

Here, it is worth pointing out that Zorn's lemma, as the most useful of all equivalents of the axiom of choice, occurs in the proofs of several theorems of crucial importance – for example, the Hans Hahn - Stefan Banach theorem in Functional Analysis, the theorem that every vector space has a basis, the Andrey Nikolayevich Tychonoff theorem in topology saying that every product of compact spaces is compact, and those theorems in abstract algebra that each ring has a maximal ideal and that each field has an algebraic closure.

1.2 Reals, Some Basic Theorems and Sequence Limits

In this section we will list the basic properties of the reals rather than to construct the reals, and then use them to derive some foundational theorems. Furthermore, we consider sequences of reals and their most important limit theorems including the Augustin Louis Cauchy criterion for the convergence of a sequence.

The properties used to define \mathbf{R} (along with an ordered pair of functions from $\mathbf{R} \times \mathbf{R}$ into \mathbf{R}) are usually regarded as axioms and classified into three different groupings.

Field Axioms. *There are two functions $+$ and \cdot on \mathbf{R} such that \mathbf{R} becomes a field:*

- F1. $x + y = y + x$ and $x \cdot y = y \cdot x$ for every $x, y \in \mathbf{R}$ (commutativity);
- F2. $(x + y) + z = x + (y + z)$ and $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for every $x, y, z \in \mathbf{R}$ (associativity);
- F3. $x \cdot (y + z) = x \cdot y + x \cdot z$ for every $x, y, z \in \mathbf{R}$ (distributivity);
- F4. *There are distinct elements 0 and 1 in \mathbf{R} such that $x + 0 = x$ and $x \cdot 1 = x$ for every $x \in \mathbf{R}$ (existence of neutrals);*
- F5. *For every $x \in \mathbf{R}$ there is an element of \mathbf{R} , denoted $-x$, such that $x + (-x) = 0$, and for every nonzero $x \in \mathbf{R}$ there is an element of \mathbf{R} , denoted $x^{-1} = 1/x$, such that $x \cdot x^{-1} = 1$ (existence of inverses).*

Here and elsewhere we will employ the common notational conventions when no confusion is possible. For example, we often write xy and $x - y$ for $x \cdot y$ and $x + (-y)$ respectively. Most of the rules of elementary algebra can be justified by the above five field axioms. For instance, for any $a, b \in \mathbf{R}$ the equation $x + a = b$ has one and only one solution $x = b - a$. The next is the group of order axioms.

Order Axioms. *There is an order relation $<$ defined on \mathbf{R} satisfying:*

- O1. *If $x, y \in \mathbf{R}$ then one and only one of: $x < y$, $x = y$, $y < x$ holds (trichotomy);*
- O2. *If $x < y$ and $y < z$ for $x, y, z \in \mathbf{R}$ then $x < z$ (transitivity);*
- O3. *If $x < y$ and $x, y, z \in \mathbf{R}$ then $x + z < y + z$;*
- O4. *If $x < y$, $x, y, z \in \mathbf{R}$ and $0 < z$ then $xz < yz$.*

We also write $y > x$ for $x < y$ and define $x \leq y$ or $y \geq x$ for either $x < y$ or $x = y$. Consequently, if $x > 0$ or $x < 0$ then we say that x is positive or negative, and according to Definition 1.1.13 (ii), (\mathbf{R}, \leq) is a totally ordered set. Naturally, the inequality $0 < 1$ follows from the above four order axioms. Furthermore, let

$$|x| = \begin{cases} x & , x > 0 \\ 0 & , x = 0 \\ -x & , x < 0 \end{cases} .$$

Then the field and order axioms can be used to derive that the triangle inequality and its reverse counterpart

$$|x + y| \leq |x| + |y| \quad \text{and} \quad ||x| - |y|| \leq |x - y|$$

hold for every $x, y \in \mathbf{R}$. The final axiom for \mathbf{R} is the following axiom which gives a further condition on the ordering of \mathbf{R} .

Axiom of Least Upper Bound (LUB). *If a subset S of \mathbf{R} is nonempty and bounded from above, then it has a least upper bound, denoted $\sup S$.*

Obviously, $\sup S$ is unique when it exists – thus it can be called the least upper bound. Using this axiom, we can prove that any nonempty subset S of \mathbf{R} having a lower bound in \mathbf{R} must have a greatest lower bound, denoted $\inf S$. In fact, a set $S \subseteq \mathbf{R}$ is bounded from below when and only when the set $T = \{x \in \mathbf{R} : -x \in S\}$ is bounded from above, and if $S \neq \emptyset$ is bounded from below then $-\sup T$ is the greatest lower bound of S .

To better understand the previously-introduced axioms, we proceed to show that the foregoing LUB-axiom is not valid for the rationals, but the reals are not

very far from the rationals in the sense that any real number may be approximated as closely as we wish. More precisely, we have the following result.

Theorem 1.2.1.

(i) If $S = \{r \in \mathbf{Q} : r > 0 \text{ and } r^2 < 2\}$, then S has an upper bound in \mathbf{Q} but does not have a least upper bound in \mathbf{Q} .

(ii) \mathbf{Q} is dense in \mathbf{R} in the sense that for $x, \varepsilon \in \mathbf{R}$ with $\varepsilon > 0$ there exists an $r \in \mathbf{Q}$ such that $|x - r| < \varepsilon$.

Proof. (i) Clearly, 2 is an upper bound of S . Suppose that S has a rational supremum s . Then by definition, for any rational ε : $0 < \varepsilon < s$, there is a rational $r \in S$ such that $0 < s - \varepsilon < r$. Consequently, $(s - r)^2 < r^2 < 2$. Note that $s + \varepsilon$ is a rational greater than s . So $s + \varepsilon \notin S$. This yields $(s + \varepsilon)^2 \geq 2$ and then $(s - \varepsilon)^2 < 2 \leq (s + \varepsilon)^2$. Now $(s - \varepsilon)^2 < s^2 \leq (s + \varepsilon)^2$. Thus

$$|s^2 - 2| < (s + \varepsilon)^2 - (s - \varepsilon)^2 = 4s\varepsilon.$$

This is valid for $0 < \varepsilon < s$, and therefore certainly for those rationals $\varepsilon \geq s$. Accordingly, $s^2 = 2$. Since s is rational, it can be written as m/n where $n \neq 0$, and $m, n \in \mathbf{Z}$ have no common factor. From this it follows that $m^2 = 2n^2$ and so that m^2 is even $2k$. Consequently, $n^2 = 2k^2$ which implies that n is even. We have seen that m and n have 2 as a common factor, contradicting that m and n have no common factor. This argument actually shows that S has no rational supremum, as desired.

(ii) Given $x, \varepsilon \in \mathbf{R}$ and $\varepsilon > 0$, in order to get a rational r obeying $|x - r| < \varepsilon$, we first verify three intermediate results as follows.

Step 1. There is an $n \in \mathbf{Z}$ such that $n > x$. To see this, assume $n \leq x$ for all $n \in \mathbf{Z}$. Then \mathbf{Z} has an upper bound x and hence it has a least upper bound in \mathbf{R} , say z , by the axiom of LUB. But $n \in \mathbf{Z}$ implies $n + 1 \in \mathbf{Z}$. So $n + 1 \leq z$, i.e., $n \leq z - 1$, thus showing that $z - 1$ is also an upper bound of \mathbf{Z} , contradicting that z is the least.

Step 2. There is an $n \in \mathbf{Z}$ such that $n \leq x < n + 1$. By *Step 1*, we choose a natural number $n_0 > |x|$, so that $-n_0 < x < n_0$. Now we take n to be the greatest element of the finite set

$$\{-n_0, -n_0 + 1, \dots, -1, 0, 1, \dots, n_0 - 1, n_0\}$$

which is less than or equal to x .

Step 3. There is an $n \in \mathbf{N}$ such that $n^{-1} < \varepsilon$. By *Step 1*, we take an integer $n > \varepsilon^{-1} > 0$, hence proving $n^{-1} < \varepsilon$ thanks to O4 above.

Now, for the above-given x, ε , we use *Step 3* to find a natural number n_0 such that $n_0^{-1} < \varepsilon$, then apply *Step 2* to n_0x to get an $n \in \mathbf{Z}$ such that $n \leq n_0x < n+1$, and hence derive $0 \leq x - nn_0^{-1} < n_0^{-1} < \varepsilon$, as desired. \square

As important consequences of the LUB-axiom of \mathbf{R} , we will prove three theorems – the nested interval theorem; the Bernard Bolzano - Karl Theodor Wilhelm Weierstrass theorem; the Eduard Heine - Félix Édouard Justin Émile Borel theorem – which will serve as useful tools in the subsequent chapters.

First of all, we make the following convention - for $a, b \in \mathbf{R}$ with $a < b$ let

$$(a, b) = \{x \in \mathbf{R} : a < x < b\} \quad \text{and} \quad [a, b] = \{x \in \mathbf{R} : a \leq x \leq b\},$$

which are called open and closed intervals with endpoints a and b respectively, and let

$$(a, b] = \{x \in \mathbf{R} : a < x \leq b\} \quad \text{and} \quad [a, b) = \{x \in \mathbf{R} : a \leq x < b\}$$

which are called open-closed or closed-open intervals with endpoints a and b respectively. In all cases, the length of an interval is equal to $b - a$.

We next introduce three concepts required respectively by the above-mentioned theorems.

Definition 1.2.2.

(i) $\{I_n\}_{n=1}^{\infty}$ is called a sequence of intervals provided that it is a function that assigns an interval I_n to each $n \in \mathbf{N}$. If $I_{n+1} \subseteq I_n$ (respectively $I_n \subseteq I_{n+1}$) for all $n \in \mathbf{N}$, then $\{I_n\}_{n=1}^{\infty}$ is called a nested downward (respectively upward) sequence of intervals.

(ii) Given $S \subseteq \mathbf{R}$ and $x \in \mathbf{R}$, x is called a cluster point of S provided that

$$(x - \varepsilon, x + \varepsilon) \cap (S \setminus \{x\}) \neq \emptyset \quad \text{for all } \varepsilon > 0.$$

(iii) A set $S \subseteq \mathbf{R}$ is said to be a closed set provided every point in S is a cluster point of S .

(iv) A set $O \subseteq \mathbf{R}$ is said to be an open set provided that its complement $O^c = \mathbf{R} \setminus O$ is closed - equivalently - any point in O is an interior point - namely - for any $x \in O$ there exists an $r > 0$ such that $(x - r, x + r) \subseteq O$.

The structure of open subsets of \mathbf{R} is described in the following Theorem 1.2.3(v) - Georg Cantor's characterization.

Theorem 1.2.3.

- (i) *The union of any two closed subsets of \mathbf{R} is closed.*
- (ii) *The intersection of any family of closed subsets of \mathbf{R} is closed.*
- (iii) *The intersection of any two open subsets of \mathbf{R} is open.*
- (iv) *The union of any family of open subsets of \mathbf{R} is open.*
- (v) *Every open subset of \mathbf{R} is the union of a countable family of pairwise disjoint open intervals.*

Proof. It is enough to verify (v). Given an open set $O \subseteq \mathbf{R}$. If $x \in O$, then there exist a point $y < x$ enjoying $(y, x) \subseteq O$, and hence $a = \inf\{y : (y, x) \subseteq O\}$ is well-defined. Correspondingly, so is $b = \sup\{z : (x, z) \subseteq O\}$. Clearly, $-\infty \leq a < b \leq \infty$. Upon writing $I(x) = (a, b) \subseteq O$, we see that $I(x)$ is an open interval containing x and $a, b \notin O$. Moreover, if $x_1 \neq x_2$ in O , then $I(x_1) = I(x_2)$ or $I(x_1) \cap I(x_2) = \emptyset$. Note that each $x \in O$ belongs to $I(x)$ which contains a rational number. So a consideration of the family of open intervals $\{I(x)\}_{x \in O}$ yields that O must be the union of a countable family of pairwise disjoint open intervals $\{I(r_j)\}_{j=1}^{n \leq \infty}$. \square

Theorem 1.2.4.

- (i) *The nested interval property: if $\{I_n\}_{n=1}^{\infty}$ is a nested downward sequence of closed intervals in \mathbf{R} , then $\cap_{n=1}^{\infty} I_n$ is a nonempty set.*
- (ii) *The Bolzano-Weierstrass property: if a subset S of \mathbf{R} is infinite and bounded from below and above, then S has at least one cluster point in \mathbf{R} .*
- (iii) *The Heine-Borel property: if \mathcal{F} is a family of open intervals covering a closed interval I in \mathbf{R} , then \mathcal{F} contains a finite subcovering; that is, there are finitely many open intervals from \mathcal{F} such that their union covers I .*

Proof. (i) Let $I_n = [a_n, b_n]$ be nested. Then $I_n \subseteq I_1$ and hence $a_n \leq b_1$ for all $n \in \mathbf{N}$. This just says that $\{a_n\}_{n=1}^{\infty}$ is bounded from above. By the axiom of LUB, the supremum $x = \sup\{a_n\}_{n=1}^{\infty}$ of $\{a_n\}_{n=1}^{\infty}$ exists in \mathbf{R} . Of course, $a_n \leq x$ for all $n \in \mathbf{N}$. Note that $\{I_j\}_{j=1}^{\infty}$ is decreasing. So it follows that

$$a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n \leq \cdots \quad \text{and} \quad b_1 \geq b_2 \geq b_3 \geq \cdots \geq b_n \geq \cdots$$

and consequently $a_m \leq a_n < b_n$ when $m \leq n$ as well as $a_m < b_m \leq b_n$ when $m \geq n$. In both case we always have $a_m \leq b_n$, thus finding that b_n is an upper bound of $\{a_n\}_{n=1}^{\infty}$ and so $x \leq b_n$. Therefore, $x \in \bigcap_{n=1}^{\infty} I_n$.

(ii) By hypothesis, S is a subset of some closed interval $I_1 = [a_1, b_1]$. Upon subdividing I_1 into two equal intervals: $[a_1, 2^{-1}(a_1 + b_1)]$; $[2^{-1}(a_1 + b_1), b_1]$, and noticing that S is infinite, we find that one of the last two intervals, denoted I_2 , must contain an infinite number of points in S . Continuing this argument produces a nested sequence of closed intervals I_n containing an infinite number of points of S . According to (i), there is a point $x \in \bigcap_{n=1}^{\infty} I_n$. This point is a limit point of S . In fact, for any $\varepsilon > 0$, there is an $n_0 \in \mathbf{N}$ such that $b_n - a_n = 2^{1-n}(b_1 - a_1) < \varepsilon$ as $n > n_0$. Since I_n contains x , it must be the case that $I_n \subseteq (x - \varepsilon, x + \varepsilon)$ when $n > n_0$. In particular,

$$I_{n_0+1} \subseteq (x - \varepsilon, x + \varepsilon) \text{ and so } (S \setminus \{x\}) \cap (x - \varepsilon, x + \varepsilon) \neq \emptyset.$$

In other words, x is a cluster point of S .

(iii) Suppose the conclusion were false, that is, \mathcal{F} has no finite subcovering for I . As in (ii), let $I_1 = I = [a_1, b_1]$. By bisecting I_1 we obtain two intervals

$$[a_1, 2^{-1}(a_1 + b_1)] \text{ and } [2^{-1}(a_1 + b_1), b_1].$$

Since I_1 can not be covered by finitely many intervals in \mathcal{F} , at least one of the last two subintervals, labelled $I_2 = [a_2, b_2]$, can not be covered by finitely many intervals in \mathcal{F} . Repeating this process we find a nested sequence of closed intervals $I_n = [a_n, b_n]$ which fail to be covered by any finite number of intervals in \mathcal{F} . An application of (i) yields a point $x \in \bigcap_{n=1}^{\infty} I_n$. Furthermore, $x \in I_1$ produces an open interval (a, b) in \mathcal{F} such that $a < x < b$. Now, if δ is the smaller of $x - a$ and $b - x$, then there is an $n_0 \in \mathbf{N}$ such that $n > n_0$ implies $b_n - a_n < \delta$. Since $x \in I_n$, it follows that $I_n \subseteq (a, b)$ as $n > n_0$. But, this contradicts the choice of I_n . Therefore, \mathcal{F} contains a finite subcover of I . □

Remark 1.2.5. *The following simple comments can help us get a better understanding of Theorem 1.2.4:*

(i) *The requirement “closed” in the nested interval property can not be replaced by “open” or “open-closed” or “closed-open”. For example, if $I_n = (0, n^{-1})$ or $(0, n^{-1}]$ or $[-n^{-1}, 0)$ then $\bigcap_{n=1}^{\infty} I_n = \emptyset$.*

(ii) Any finite set has no cluster point. Of course, an infinite set may fail to have cluster points – for example, \mathbf{N} . Nevertheless, according to the Bolzano-Weierstrass theorem, any bounded (from above and below) infinite subset of \mathbf{R} must have at least one cluster point.

(iii) Although $(0, 1) \subseteq \bigcup_{n=3}^{\infty} (n^{-1}, 1 - n^{-1})$, the open interval $(0, 1)$ cannot be covered by any finitely many intervals from $\{(n^{-1}, 1 - n^{-1})\}_{n=3}^{\infty}$. However, this does not violate the Heine-Borel property which is valid for all finite closed intervals.

We close this section by discussing sequences of reals and their limits.

Definition 1.2.6.

- (i) A sequence $\{s_n\}_{n=1}^{\infty}$ in \mathbf{R} is a function from \mathbf{N} into \mathbf{R} . We use $\{s_n\}_{n=1}^{\infty} = \{s_n : n \in \mathbf{N}\}$ as the range of the sequence.
- (ii) A sequence $\{s_n\}_{n=1}^{\infty}$ is called bounded provided that its range is bounded, i.e., there exists a $b > 0$ such that $\sup\{|s_n|\}_{n=1}^{\infty} \leq b$.
- (iii) For $s \in \mathbf{R}$ and a sequence $\{s_n\}_{n=1}^{\infty}$, we say that $\{s_n\}_{n=1}^{\infty}$ converges (or is convergent) to s , denoted either $\lim_{n \rightarrow \infty} s_n = s$ or $s_n \rightarrow s$, if and only if for any $\varepsilon > 0$ there exists an $n_0 \in \mathbf{N}$ such that $n > n_0$ implies $|s_n - s| < \varepsilon$. A sequence that is not convergent is called divergent.

It is easy to see that a convergent sequence has only one limit, but also is bounded. Moreover, if $\lim_{n \rightarrow \infty} s_n = s \neq 0$, then for $\varepsilon = 2^{-1}|s|$ there is an $n_0 \in \mathbf{N}$ such that $|s_n - s| < \varepsilon$ as $n > n_0$, and hence $|s_n| \geq |s| - \varepsilon = 2^{-1}|s|$. Using these facts and the triangle inequality, we can establish the following limit theorem.

Theorem 1.2.7. For $s, t \in \mathbf{R}$ and sequences $\{s_n\}_{n=1}^{\infty}$ and $\{t_n\}_{n=1}^{\infty}$ in \mathbf{R} , let $\lim_{n \rightarrow \infty} s_n = s$ and $\lim_{n \rightarrow \infty} t_n = t$. Then:

- (i) $\lim_{n \rightarrow \infty} (s_n + t_n) = s + t$;
- (ii) $\lim_{n \rightarrow \infty} (s_n t_n) = st$;
- (iii) $\lim_{n \rightarrow \infty} (cs_n) = cs$ for any $c \in \mathbf{R}$;
- (iv) $\lim_{n \rightarrow \infty} (s_n t_n^{-1}) = st^{-1}$ provided that $t \neq 0$ and $t_n \neq 0$ for each $n \in \mathbf{N}$;
- (v) $s \leq t$ provided that $s_n \leq t_n$ for each $n \in \mathbf{N}$.

Proof. We leave the demonstration as an exercise. □

Definition 1.2.8.

(i) A function f from \mathbf{N} into \mathbf{N} is called increasing provided $f(m) \leq f(n)$ for $m, n \in \mathbf{N}$ with $m < n$. Moreover, a sequence $\{x_n\}_{n=1}^{\infty}$ in \mathbf{R} is called a subsequence of a given sequence $\{s_n\}_{n=1}^{\infty}$ in \mathbf{R} when there is an increasing function $f : \mathbf{N} \rightarrow \mathbf{N}$ such that $x_n = s_{f(n)}$ for each $n \in \mathbf{N}$.

(ii) A sequence $\{s_n\}_{n=1}^{\infty}$ in \mathbf{R} is called increasing (respectively, strictly increasing) if $s_n \leq s_{n+1}$ (respectively, $s_n < s_{n+1}$) for any $n \in \mathbf{N}$. Meanwhile, a sequence $\{s_n\}_{n=1}^{\infty}$ in \mathbf{R} is called decreasing (respectively, strictly decreasing) if $s_n \geq s_{n+1}$ (respectively, $s_n > s_{n+1}$) for each $n \in \mathbf{N}$. Furthermore, a sequence is called monotonic provided it is either increasing or decreasing.

The following result, in which (i) characterizes the convergence of a sequence by its subsequences and (ii) is the monotone convergence theorem, is natural, important and useful.

Theorem 1.2.9. *Let $\{s_n\}_{n=1}^{\infty}$ be a sequence in \mathbf{R} . Then:*

- (i) $\{s_n\}_{n=1}^{\infty}$ is convergent to $s \in \mathbf{R}$ when and only when each subsequence of $\{s_n\}_{n=1}^{\infty}$ is convergent to $s \in \mathbf{R}$;
- (ii) $\{s_n\}_{n=1}^{\infty}$ is convergent whenever it is monotonic and bounded.

Proof. (i) Suppose that $s_n \rightarrow s$ and $\{s_{n_k}\}_{k=1}^{\infty}$ is a subsequence of $\{s_n\}_{n=1}^{\infty}$. Then for $\varepsilon > 0$ there is an $n_0 \in \mathbf{N}$ such that $n > n_0$ implies $|s_n - s| < \varepsilon$, and consequently, if $k > n_0$ then $n_k \geq k > n_0$ and hence $|s_{n_k} - s| < \varepsilon$, that is, $s_{n_k} \rightarrow s$. Conversely, suppose that every subsequence of $\{s_n\}_{n=1}^{\infty}$ is convergent to s . If $\{s_n\}_{n=1}^{\infty}$ is not convergent to $s \in \mathbf{R}$, then by negation of the limit definition, there is $\varepsilon_0 > 0$ such that for any $k \in \mathbf{N}$ one can always find s_{n_k} from $\{s_n\}_{n=1}^{\infty}$ such that $|s_{n_k} - s| \geq \varepsilon_0$. This actually means that $\{s_{n_k}\}_{k=1}^{\infty}$ does not converge to s , a contradiction.

(ii) It suffices to consider the case that $\{s_n\}_{n=1}^{\infty}$ is increasing and bounded since the other case can be handled similarly. To this end, we see from the axiom of LUB that $a = \sup\{s_n\}_{n=1}^{\infty}$ is a finite element of \mathbf{R} . Also, by definition, for every $\varepsilon > 0$ there is an $n_0 \in \mathbf{N}$ such that $a - \varepsilon < s_{n_0}$. Hence $a - \varepsilon < s_{n_0} \leq s_n \leq a$ as $n > n_0$. Consequently, $s_n \rightarrow a$. \square

Without using subsequences we can also characterize the convergence of a sequence. To see this, we introduce the notion of Cauchy's sequence.

Definition 1.2.10. *A sequence $\{s_n\}_{n=1}^{\infty}$ in \mathbf{R} is called a Cauchy sequence provided that for every $\varepsilon > 0$ there exists an $n_0 \in \mathbf{N}$ such that $|s_m - s_n| < \varepsilon$ for $m, n > n_0$.*

The next property is the so-called completeness of \mathbf{R} .

Theorem 1.2.11. *A sequence in \mathbf{R} is convergent if and only if it is a Cauchy sequence.*

Proof. Suppose that the sequence $\{s_n\}_{n=1}^{\infty}$ in \mathbf{R} is convergent to s . Then for any $\varepsilon > 0$ there is $n_0 \in \mathbf{N}$ such that $n > n_0$ implies $|s_n - s| < 2^{-1}\varepsilon$ and consequently, if $m, n > n_0$ then by the triangle inequality,

$$|s_m - s_n| \leq |s_m - s| + |s_n - s| < \varepsilon,$$

namely, $\{s_n\}_{n=1}^{\infty}$ is a Cauchy sequence.

Conversely, if $\{s_n\}_{n=1}^{\infty}$ is a Cauchy sequence, then there is an $n_0 \in \mathbf{N}$ such that $|s_{n+1} - s_{n_0+1}| < 1$ and thus $|s_{n+1}| < 1 + |s_{n_0+1}|$ whenever $n > n_0$. Accordingly, $\{s_n\}_{n=1}^{\infty}$ is bounded. Let us consider two cases as follows.

Case 1: $\{s_n\}_{n=1}^{\infty}$ is finite. This case yields that some member, say s in this sequence, must be repeated infinitely. Thus, there exists a subsequence $\{s_{n_k}\}_{k=1}^{\infty}$ such that $|s_{n_k} - s| = 0$ for all $k \in \mathbf{N}$.

Case 2: $\{s_n\}_{n=1}^{\infty}$ is infinite. Then using the Bolzano-Weierstrass property, we can also get a subsequence $\{s_{n_k}\}_{k=1}^{\infty}$ that is convergent to some $s \in \mathbf{R}$. Consequently, for any $\varepsilon > 0$ there are $k_1, k_2 \in \mathbf{N}$ such that $|s_{n_k} - s| < 2^{-1}\varepsilon$ as $k > k_1$ and $|s_m - s_n| < 2^{-1}\varepsilon$ for $m, n > k_2$. Finally, if $n > k_3 = \max\{k_1, k_2\}$ then by taking $k \geq k_3$ we derive

$$|s_n - s| \leq |s_n - s_{n_k}| + |s_{n_k} - s| < 2^{-1}\varepsilon + 2^{-1}\varepsilon = \varepsilon,$$

thereby proving $s_n \rightarrow s$. □

Remark 1.2.12.

(i) *The argument for Theorem 1.2.11 yields the Bolzano-Weierstrass property for sequences - any bounded sequence in \mathbf{R} has a convergent subsequence.*

(ii) *Also, Theorem 1.2.11 produces another way to look at the difference between \mathbf{R} and \mathbf{Q} . To be more specific, let $\{s_n\}_{n=1}^{\infty}$ be a sequence of rationals with $s_n \rightarrow \sqrt{2}$ (by the density of \mathbf{Q} in \mathbf{R}). Then this sequence is a Cauchy sequence in \mathbf{Q} but its limit $\sqrt{2}$ is in $\mathbf{R} \setminus \mathbf{Q}$. In other words, \mathbf{Q} is not complete, but \mathbf{R} is.*

Problems

1.1. The symmetric difference of two given sets X and Y is defined as $(X \cup Y) \setminus (X \cap Y)$. Prove $(X \cup Y) \setminus (X \cap Y) = (X \setminus Y) \cup (Y \setminus X)$.

1.2. For $j \in \mathbf{N}$ let $X_j = [0, j^{-1}]$. Find $\bigcup_{j=1}^{\infty} X_j$ and $\bigcap_{j=1}^{\infty} X_j$.

1.3. Given a set X , let $\{X_j\}_{j \in I}$ be a family of sets indexed by I . Prove

$$X \cap (\bigcup_{j \in I} X_j) = \bigcup_{j \in I} (X \cap X_j) \quad \text{and} \quad X \cup (\bigcap_{j \in I} X_j) = \bigcap_{j \in I} (X \cup X_j).$$

1.4. Prove that if X , Y and Z are sets then $X \times (Y \cup Z) = (X \times Y) \cup (X \times Z)$.

1.5. Let (a, b) be an open interval of \mathbf{R} . Find a bijection from $(0, 1)$ onto (a, b) and then onto $(0, \infty)$.

1.6. Prove that if $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ are two functions and C is a subset of Z then $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$.

1.7. Prove that if X is an infinite set and $x \in X$ then $X \sim_{\mathbf{R}} X \setminus \{x\}$.

1.8. Suppose that f is a one-to-one function from X into Y with Y countable. Show that X is countable.

1.9. Prove that if X is the set of all functions $f : [0, 1] \rightarrow \mathbf{R}$ then there is no function from $[0, 1]$ onto X .

1.10. Use the field axioms to show the following properties.

(i) If $a, b \in \mathbf{R}$ with $a \neq 0$ then the equation $ax = b$ has one and only one solution.

(ii) If $a \in \mathbf{R}$ is nonzero then $(a^{-1})^{-1} = a$.

(iii) If $a, b \in \mathbf{R}$ are nonzero then $(ab)^{-1} = a^{-1}b^{-1}$.

1.11. Use the mathematical induction to prove that

$$|a_1 + a_2 + \cdots + a_n| \leq |a_1| + |a_2| + \cdots + |a_n|$$

is valid for $a_1, a_2, \dots, a_n \in \mathbf{R}$.

1.12. Let $a, b \geq 0$. Prove that $a \leq b$ if and only if $a^3 \leq b^3$.

1.13. For two nonempty subsets X and Y of \mathbf{R} let $X + Y = \{x + y : x \in X \text{ and } y \in Y\}$. Prove

$$\sup(X + Y) = \sup X + \sup Y \quad \text{and} \quad \inf(X + Y) = \inf X + \inf Y.$$

1.14. In the notation of Theorem 1.2.4 and its proof, verify the following three facts.

(i) $a = \lim_{n \rightarrow \infty} a_n$ and $b = \lim_{n \rightarrow \infty} b_n$ are in $\cap_{n=1}^{\infty} I_n$.

(ii) $[a, b] = \cap_{n=1}^{\infty} I_n$.

(iii) $\cap_{n=1}^{\infty} I_n$ is a single point provided that $\lim_{n \rightarrow \infty} (b_n - a_n) = 0$.

1.15. Prove Theorem 1.2.7.

1.16. A sequence $\{s_n\}_{n=1}^{\infty}$ is said to diverge to ∞ respectively $-\infty$, denoted $\lim_{n \rightarrow \infty} s_n = \infty$ respectively $\lim_{n \rightarrow \infty} s_n = -\infty$, provided that for any $a \in \mathbf{R}$ there is an $n_0 \in \mathbf{N}$ such that $n > n_0$ implies $s_n > a$ respectively $s_n < a$. Prove that if $\{s_n\}_{n=1}^{\infty}$ is a sequence of positive numbers in \mathbf{R} then $\lim_{n \rightarrow \infty} s_n = \infty$ if and only if $\lim_{n \rightarrow \infty} s_n^{-1} = 0$.

1.17. For $n \in \mathbf{N}$ let $s_n = 1 + 2^{-1} + \cdots + n^{-1}$. Prove that:

(i) $\{s_n\}_{n=1}^{\infty}$ is increasing but divergent;

(ii) $\{s_n\}_{n=1}^{\infty}$ is not a Cauchy sequence although $\lim_{n \rightarrow \infty} |x_{n+1} - x_n| = 0$.

1.18. Let $\{s_n\}_{n=1}^{\infty}$ be a bounded sequence in \mathbf{R} . A subsequential limit of $\{s_n\}_{n=1}^{\infty}$ is a real number that is the limit of some subsequence of $\{s_n\}_{n=1}^{\infty}$. If S is the set of all subsequential limits of $\{s_n\}_{n=1}^{\infty}$, then we define the limit superior/upper limit and the limit inferior/lower limit of $\{s_n\}_{n=1}^{\infty}$ to be

$$\limsup s_n = \sup S \quad \text{and} \quad \liminf s_n = \inf S.$$

Prove the following two facts.

(i)

$$\limsup s_n = \lim_{m \rightarrow \infty} \sup \{s_n : n > m\} \quad \text{and} \quad \liminf s_n = \lim_{m \rightarrow \infty} \inf \{s_n : n > m\}.$$

(ii) $\{s_n\}_{n=1}^{\infty}$ is convergent if and only if $\limsup s_n = \liminf s_n$.

Chapter 2

Riemann Integrals

The Riemann integral – the first rigorous definition of the integral of a function on an interval, created by Bernhard Riemann, is the integral normally encountered in Real Analysis and used by physicists and engineers. In this chapter, we discuss the definition and basic properties of the Riemann integral for real-valued functions of one real variable. Here we are concerned with the simplest results, up to Riemann sums, the integrability of a continuous function and the fundamental theorem of calculus. Moreover, we consider the improper integral as the limit of Riemann’s integral on an interval, when an endpoint, or both endpoints, of the interval approaches either a specified real number or infinity.

2.1 Definitions, Examples and Basic Properties

From now on, we denote the implication “if A then B ” by $A \Rightarrow B$ whenever appropriate. Without a special remark, $[a, b]$ always stands for a finite closed interval in \mathbf{R} . Let f be a real-valued function on the interval $[a, b]$. A partition P of $[a, b]$ is a set $\{x_k\}_{k=0}^n$ satisfying $a = x_0 < x_1 < \cdots < x_n = b$, $n \in \mathbf{N}$. The length of the subinterval $[x_{k-1}, x_k]$ is written as $\Delta x_k = x_k - x_{k-1}$, and the norm $\|P\|$ of the partition P is defined by

$$\|P\| = \max \{\Delta x_k : k = 1, 2, \dots, n\}.$$

Thus P determines n subintervals of $[a, b]$ of which the largest has length $\|P\|$, and the subinterval $[x_{k-1}, x_k]$ is called the k -th subinterval associated with P . In each of the n subintervals, choose a number $\xi_k \in [x_{k-1}, x_k]$, and form the

following sum

$$S(f, P, \xi) = S(f, P, \{\xi_k\}_{k=1}^n) = \sum_{k=1}^n f(\xi_k) \Delta x_k.$$

This is called a Riemann sum for the function f on $[a, b]$. With some familiarity with integral calculus, we can immediately recognize that if f is nonnegative then $S(f, P, \xi)$ gives an approximation to the area of the region between the graph of f and the horizontal axis. Needless to say, if f is not continuous on $[a, b]$, then it may be hard to think of the graph of f as serving as a boundary for any such region. Another interpretation which has the advantage of being free of geometrical dependence, is that $S(f, P, \xi)$ represents a certain average value of f on $[a, b]$. This can be seen as follows:

$$(b-a)^{-1} S(f, P, \xi) = \sum_{k=1}^n f(\xi_k) (b-a)^{-1} \Delta x_k.$$

In the sum, $f(\xi_k)$ is multiplied by $(b-a)^{-1} \Delta x_k$ that gives the fractional part of $[a, b]$ from which ξ_k is chosen. Thus the sum is a weighted average of the n values $\{f(\xi_k)\}_{k=1}^n$. Multiplying this weighted average by the length $b-a$ of $[a, b]$, we get $S(f, P, \xi)$. Therefore it is natural to expect that $(b-a)^{-1} S(f, P, \xi)$ gives an average of the values of f on $[a, b]$. This average is a better indicator of the behavior of f if P determines only small subintervals, so we examine the behavior of $S(f, P, \xi)$ as $\|P\|$ tends to 0.

Definition 2.1.1. f is called Riemann integrable on $[a, b]$ provided $\lim_{\|P\| \rightarrow 0} S(f, P, \xi)$ exists. In this case the limit value, denoted $\int_a^b f(x) dx$, is said to be the Riemann integral of f on $[a, b]$. The set of all Riemann integrable functions on $[a, b]$ is written as $R[a, b]$.

Remark 2.1.2. We must acknowledge that the limit $\lim_{\|P\| \rightarrow 0} S(f, P, \xi)$ in Definition 2.1.1 is neither a sequential limit nor a function limit (because $S(f, P, \xi)$ is not a function of $\|P\|$). In order to fully appreciate the latter assertion, we consider the fact that for a given value of $\|P\|$, there could be many partitions of $[a, b]$ whose largest subinterval has length $\|P\|$. For example, $[0, 1]$ may be partitioned by

$$P_1 = \{0, 3^{-1}, 2 \cdot 3^{-1}, 1\}; P_2 = \{0, 3^{-1}, 2^{-1}, 3 \cdot 2^{-2}, 1\}; P_3 = \{0, 5^{-1}, 2 \cdot 5^{-1}, 2^{-1}, 2 \cdot 3^{-1}, 1\},$$

and then

$$\|P_1\| = \|P_2\| = \|P_3\| = 3^{-1}.$$

Moreover, for each partition of $[a, b]$ having the given norm $\|P\|$, there are many ways of choosing $\{\xi_k\}_{k=1}^n$ from the n subintervals. Thus the value of $\|P\|$ does not determine $S(f, P, \xi)$. It is therefore necessary to give a complete description of the limit concept used in Definition 2.1.1.

Definition 2.1.3. That $\lim_{\|P\| \rightarrow 0} S(f, P, \xi) = s \in \mathbf{R}$ exists means that if $\varepsilon > 0$ then there exists a $\delta > 0$ such that for every partition P of $[a, b]$ with $\|P\| < \delta$ and any choice of the points $\{\xi_k \in [x_{k-1}, x_k]\}_{k=1}^n$, the inequality $|S(f, P, \xi) - s| < \varepsilon$ holds.

Usually, a direct argument for the existence of $\lim_{\|P\| \rightarrow 0} S(f, P, \xi)$ is too complicated to handle, but Definition 2.1.3 for $\int_a^b f(x) dx$ has certain advantages for developing the theory. In the forthcoming section we prove an equivalent formulation of the limit which is easier to use in proving the integrability of particular functions. Meanwhile, however, we deal first with several examples.

Example 2.1.4. If f is a constant κ on $[a, b]$, then $f \in R[a, b]$ and $\int_a^b f(x) dx = \kappa(b - a)$.

In fact, for any partition P of $[a, b]$, we have

$$S(f, P, \xi) = \kappa \sum_{k=1}^n \Delta x_k = \kappa(x_n - x_0) = \kappa(b - a).$$

Thus, for a given $\varepsilon > 0$, the inequality $|S(f, P, \xi) - \kappa(b - a)| = 0 < \varepsilon$ is satisfied trivially. Hence

$$\int_a^b f(x) dx = \lim_{\|P\| \rightarrow 0} S(f, P, \xi) = \kappa(b - a).$$

Example 2.1.5. Let c be a fixed point of $[a, b]$ and

$$f(x) = \begin{cases} 1, & x = c \\ 0, & x \neq c \end{cases},$$

then $f \in R[a, b]$ and $\int_a^b f(x) dx = 0$.

As a matter of fact, for any $S(f, P, \xi)$ corresponding to a partition P of $[a, b]$ we have $|S(f, P, \xi)| \leq 2\|P\|$, where 2 appears since c may be one of the partition points x_k and we may in this case have both ξ_k and ξ_{k+1} equal to c . So $\int_a^b f(x) dx = 0$.

Example 2.1.6. Let $[\alpha, \beta] \subset [a, b]$ and

$$f(x) = \begin{cases} 1, & x \in (\alpha, \beta) \\ 0, & x \in [a, b] \setminus (\alpha, \beta) \end{cases},$$

then $f \in R[a, b]$ and $\int_a^b f(x) dx = \beta - \alpha$.

To verify this, suppose that $P = \{x_0, x_1, \dots, x_n\}$ is a partition of $[a, b]$ obeying $\|P\| < \delta$, and $S(f, P, \xi)$ corresponds to this P . Since $f(\xi_k)$ is 1 or 0 according as $\xi_k \in (\alpha, \beta)$ or not, we have

$$S(f, P, \xi) = \sum_{\xi_k \in (\alpha, \beta)} \Delta x_k.$$

Now choose l, j from $\{1, 2, \dots, n\}$ such that

$$x_{l-1} \leq \alpha < x_l \quad \text{and} \quad x_{j-1} \leq \beta < x_j.$$

Then $\xi_k \in (\alpha, \beta)$ whenever $l+1 \leq k \leq j-1$, as well as, $\xi_k \notin (\alpha, \beta)$ whenever $k < l$ or $k > j$. Therefore

$$\sum_{l+1 \leq k \leq j-1} \Delta x_k \leq S(f, P, \xi) \leq \sum_{l \leq k \leq j} \Delta x_k.$$

By the choice of l and j , we see

$$\beta - \alpha \leq x_j - x_{l-1} < (j - l + 1)\|P\|,$$

so that if δ is sufficiently small then $l+1 \leq j-1$ must be true, and hence

$$x_{j-1} - x_l \leq S(f, P, \xi) \leq x_j - x_{l-1}.$$

Hence

$$(x_{j-1} - \beta) - (x_l - \alpha) \leq S(f, P, \xi) - (\beta - \alpha) \leq (x_j - \beta) - (x_{l-1} - \alpha).$$

Note that $x_{j-1} - \beta$, $x_l - \alpha$, $x_j - \beta$ and $x_{l-1} - \alpha$ are of absolute values less than $\|P\|$. So

$$|S(f, P, \xi) - (\beta - \alpha)| < 2\delta.$$

This implies that $\int_a^b f(x) dx = \beta - \alpha$.

Example 2.1.7. *If*

$$f(x) = \begin{cases} 1, & x \in \mathbf{Q} \\ 0, & x \notin \mathbf{Q} \end{cases},$$

then $f \notin R[a, b]$ for any interval $[a, b]$.

To prove the result, we take $\varepsilon = 2^{-1}(b - a)$ and use Theorem 1.2.1 (ii) – the density of the rational numbers in \mathbf{R} , equivalently, the density of the irrational numbers in \mathbf{R} . For any partition P of $[a, b]$, no matter how small $\|P\|$ may be, each subinterval associated with P must contain both rational numbers and irrational numbers, say,

$$\hat{\xi}_k \in [x_{k-1}, x_k] \cap \mathbf{Q} \text{ and } \check{\xi}_k \in [x_{k-1}, x_k] \cap (\mathbf{R} \setminus \mathbf{Q}).$$

Then

$$\begin{cases} S(f, P, \hat{\xi}) = \sum_{k=1}^n f(\hat{\xi}_k) \Delta x_k = b - a; \\ S(f, P, \check{\xi}) = \sum_{k=1}^n f(\check{\xi}_k) \Delta x_k = 0. \end{cases}$$

Since these two numbers are $b - a$ units apart, there is no such number I that can be within ε of both sums. Hence, $\lim_{\|P\| \rightarrow 0} S(f, P, \xi)$ does not exist, so f is not in $R[a, b]$.

Because of working on an entirely new limit concept, we must begin by proving such basic results as the uniqueness of the limit.

Theorem 2.1.8. *If $f \in R[a, b]$, then $\int_a^b f(x) dx$ is unique and f is bounded.*

Proof. Suppose that $\lim_{\|P\| \rightarrow 0} S(f, P, \xi) = s$, let $t \neq s$ and take $\varepsilon = 2^{-1}|s - t|$. Since ε is half of the distance between s and t , it follows that $(s - \varepsilon, s + \varepsilon)$ and $(t - \varepsilon, t + \varepsilon)$ do not intersect. When P is a partition of $[a, b]$ with $\|P\|$ being sufficiently small, $S(f, P, \xi)$ is in $(s - \varepsilon, s + \varepsilon)$ for any choice of $\{\xi_k\}_{k=1}^n$. Then $S(f, P, \xi)$ is not in $(t - \varepsilon, t + \varepsilon)$, so $|S(f, P, \xi) - t| \geq \varepsilon$. Hence $\lim_{\|P\| \rightarrow 0} S(f, P, \xi)$ cannot equal t .

Let us next prove the boundedness of f on $[a, b]$. If f is not bounded on $[a, b]$, then for any partition P of $[a, b]$, f must be unbounded on at least one of the subintervals associated with P , say, $[x_{m-1}, x_m]$. For $k \neq m$, choose $\{\xi_k\}_{k=1}^n$ satisfying $x_{k-1} \leq \xi_k \leq x_k$. Thus for any $s \in \mathbf{R}$ we have determined the number $|s| + \sum_{k \neq m} |f(\xi_k)| \Delta x_k$. Now we use the unboundedness of f on $[x_{m-1}, x_m]$ to choose ξ_m so that

$$|f(\xi_m)(x_m - x_{m-1})| > 1 + |s| + \sum_{k \neq m} |f(\xi_k)| \Delta x_k.$$

Thus the m -th term dominates $S(f, P, \xi)$, and we have $|S(f, P, \xi)| > 1 + |s|$. Hence $\lim_{\|P\| \rightarrow 0} S(f, P, \xi)$ cannot exist, so f does not belong to $R[a, b]$, a contradiction. \square

The next theorem shows that an integral can be evaluated on two adjacent intervals and the resulting values are added.

Theorem 2.1.9. *If $f \in R[a, b]$ and $f \in R[b, c]$, then $f \in R[a, c]$ and $\int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx$.*

Proof. Any partition P of $[a, c]$ determines partitions P_1 and P_2 on $[a, b]$ and $[b, c]$, respectively, since if b is not one of the partition points of P then it can be inserted between x_{j-1} and x_j where j is the least integer such that $b \leq x_j$. The number ξ_j in the sum $S(f, P, \xi)$ is selected from the subinterval $[x_{j-1}, x_j]$, and we may have either $\xi_j \in [x_{j-1}, b]$ or $\xi_j \in [b, x_j]$. In the former case, we can write

$$\begin{aligned} S(f, P, \xi) &= \sum_{k=1}^{j-1} f(\xi_k) \Delta x_k \\ &\quad + f(\xi_j)(b - x_{j-1}) - f(b)(x_j - b) + f(\xi_j)(x_j - b) + f(b)(x_j - b) \\ &\quad + \sum_{k=j+1}^n f(\xi_k) \Delta x_k \\ &= S(f, P_1, \hat{\xi}) + (f(\xi_j) - f(b))(x_j - b) + S(f, P_2, \check{\xi}), \end{aligned}$$

where $\hat{\xi} = \{\xi_k\}_{k=1}^j$ and $\check{\xi} = \{b\} \cup \{\xi_k\}_{k=j+1}^n$. Similarly, if $\xi_j \in [b, x_j]$, then

$$S(f, P, \xi) = S(f, P_1, \hat{\xi}) + (f(\xi_j) - f(b))(b - x_{j-1}) + S(f, P_2, \check{\xi}),$$

where $\hat{\xi} = \{\xi_k\}_{k=1}^{j-1} \cup \{b\}$ and $\check{\xi} = \{\xi_k\}_{k=j}^n$. By the boundedness of f on $[a, c]$, we have a constant $\kappa > 0$ such that $|f(x)| < \kappa$ on $[a, c]$ and then

$$|S(f, P, \xi) - S(f, P_1, \hat{\xi}) - S(f, P_2, \check{\xi})| \leq 2\kappa \|P\|,$$

due to

$$b - x_{j-1} \leq x_j - x_{j-1} \leq \|P\| \quad \text{and} \quad x_j - b \leq x_j - x_{j-1} \leq \|P\|.$$

If $\varepsilon > 0$, then $f \in R[a, b] \cap R[b, c]$ allows us to choose a $\delta > 0$ so that $\max\{\|P_1\|, \|P_2\|\} < \delta$ implies

$$\left| S(f, P_1, \hat{\xi}) - \int_a^b f(x) dx \right| < 3^{-1}\varepsilon \quad \text{and} \quad \left| S(f, P_2, \check{\xi}) - \int_b^c f(x) dx \right| < 3^{-1}\varepsilon.$$

We can also choose δ even smaller, if necessary, so that $\delta < \varepsilon/(6\kappa)$. Now since $\max\{\|P_1\|, \|P_2\|\} \leq \|P\|$, we conclude that $\|P\| < \delta$ implies

$$\begin{aligned}
 \left| S(f, P, \xi) - \left(\int_a^b f(x) dx + \int_b^c f(x) dx \right) \right| & \\
 \leq \left| S(f, P, \xi) - S(f, P_1, \hat{\xi}) - S(f, P_2, \check{\xi}) \right| & \\
 + \left| S(f, P_1, \hat{\xi}) - \int_a^b f(x) dx \right| + \left| S(f, P_2, \check{\xi}) - \int_b^c f(x) dx \right| & \\
 < 2\kappa\|P\| + 3^{-1}\varepsilon + 3^{-1}\varepsilon & \\
 < \varepsilon. &
 \end{aligned}$$

Hence

$$\int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx.$$

□

From Theorem 2.1.9 it follows that

$$\int_a^a f(x) dx = 0 \quad \text{and} \quad \int_a^b f(x) dx = - \int_b^a f(x) dx$$

should be set. Moreover, motivated by the argument of Theorem 2.1.9, we find that in averaging a large set of numbers, it is possible to change a few values without altering the average very much. This is true for $S(f, P, \xi)$ and the resulting integral value.

Theorem 2.1.10. *If $f \in R[a, b]$ and $g = f$ for all but a finite number of points in $[a, b]$, then $g \in R[a, b]$ and $\int_a^b g(x) dx = \int_a^b f(x) dx$.*

Proof. It suffices to prove that the assertion in which g differs from f at exactly one point in $[a, b]$, because we can produce finite changes in f by changing the value at one point and repeating the procedure finitely many times. So we may assume that $g(x) = f(x)$ for all $x \in [a, b] \setminus \{c\}$. For any partition P , the point c can be in at most two of the subintervals in that c may be a partition point, say,

$[x_{m-1}, x_m]$ and $[x_m, x_{m+1}]$. Then

$$\begin{aligned}
 & \left| S(g, P, \xi) - \int_a^b f(x) dx \right| \\
 & \leq \left| \sum_{k=1}^n (g(\xi_k) - f(\xi_k)) \Delta x_k \right| + \left| S(f, P, \xi) - \int_a^b f(x) dx \right| \\
 & \leq |g(c) - f(c)|(x_m - x_{m-1}) + |g(c) - f(c)|(x_{m+1} - x_m) \\
 & \quad + \left| S(f, P, \xi) - \int_a^b f(x) dx \right| \\
 & \leq |g(c) - f(c)|(2\|P\|) + \left| S(f, P, \xi) - \int_a^b f(x) dx \right|.
 \end{aligned}$$

Given $\varepsilon > 0$, let $\delta \leq \varepsilon/(4|g(c) - f(c)|)$. So, for any partition P of $[a, b]$ we have

$$\|P\| < \delta \Rightarrow \left| S(f, P, \xi) - \int_a^b f(x) dx \right| < 2^{-1}\varepsilon.$$

Upon substituting this into

$$|g(c) - f(c)|(2\|P\|) + \left| S(f, P, \xi) - \int_a^b f(x) dx \right|,$$

we see

$$\|P\| < \delta \Rightarrow \left| S(g, P, \xi) - \int_a^b f(x) dx \right| < |g(c) - f(c)|(2\delta) + 2^{-1}\varepsilon < \varepsilon.$$

□

Here it is worthwhile pointing out that the finiteness in Theorem 2.1.10 is important; see Example 2.1.7.

2.2 Algebraic Operations and the Darboux Criterion

Theorem 2.2.1. *If $f_1, f_2 \in R[a, b]$ and $c_1, c_2 \in \mathbf{R}$, then $c_1 f_1 + c_2 f_2 \in R[a, b]$.*

Proof. It suffices to prove $f_1 + f_2 \in R[a, b]$. Assume $\varepsilon > 0$ and choose such $\delta_1 > 0$ and $\delta_2 > 0$ that

$$\|P\| < \delta_1 \Rightarrow \left| S(f_1, P, \xi) - \int_a^b f_1(x) dx \right| < 2^{-1}\varepsilon$$

and

$$\|P\| < \delta_2 \Rightarrow \left| S(f_2, P, \xi) - \int_a^b f_2(x) dx \right| < 2^{-1}\epsilon.$$

Now, defining $\delta = \min\{\delta_1, \delta_2\}$ guarantees that: if $\|P\| < \delta$, then

$$\begin{aligned} \left| S(f_1 + f_2, P, \xi) - \left(\int_a^b f_1(x) dx + \int_a^b f_2(x) dx \right) \right| \\ \leq \left| S(f_1, P, \xi) - \int_a^b f_1(x) dx \right| + \left| S(f_2, P, \xi) - \int_a^b f_2(x) dx \right| < \epsilon. \end{aligned}$$

and hence

$$f_1 + f_2 \in R[a, b] \text{ and } \int_a^b (f_1(x) + f_2(x)) dx = \int_a^b f_1(x) dx + \int_a^b f_2(x) dx,$$

□

When evaluating a limit of some expression that is bounded above or below, we may use the same bound to control the limit value.

Corollary 2.2.2. *If $f_1, f_2 \in R[a, b]$ and $f_1(x) \leq f_2(x)$ for all $x \in [a, b]$, then $\int_a^b f_1(x) dx \leq \int_a^b f_2(x) dx$.*

Proof. Consider the case in which f_1 is identically 0. Then $f_2 \geq 0$ on $[a, b]$, so $S(f_2, P, \xi) \geq 0$. If $s_2 = \int_a^b f_2(x) dx < 0$ and $\epsilon = -s_2$, then it is impossible to have $|S(f_2, P, \xi) - s_2| < \epsilon$ because the distance between $S(f_2, P, \xi)$ and s_2 is at least equal to $|s_2|$.

Now for the general case: $f_1 \leq f_2$ on $[a, b]$. Define $f = f_2 - f_1$. Clearly, $f \geq 0$, and, by Theorem 2.2.1, $f \in R[a, b]$. So the first case of this argument ensures

$$0 \leq \int_a^b f(x) dx = \int_a^b f_2(x) dx - \int_a^b f_1(x) dx.$$

□

In order to prove that multiplication is closed in $R[a, b]$, we need to develop a characterization of $R[a, b]$ (the so-called Jean Gaston Darboux's criterion for integrability) that does not rely on predicting the value of the limit $\lim_{\|P\| \rightarrow 0} S(f, P, \xi)$. In this sense, it is analogous to Cauchy's criterion for sequential convergence.

Let f be a bounded function on $[a, b]$ and P be a partition of $[a, b]$. Due to the boundedness of f on each $[x_{k-1}, x_k]$, we may define

$$m_k = \inf\{f(x) : x \in [x_{k-1}, x_k]\} \text{ and } M_k = \sup\{f(x) : x \in [x_{k-1}, x_k]\},$$

and form the lower sum and upper sum for f with respect to P :

$$L(f, P) = \sum_{k=1}^n m_k \Delta x_k \text{ and } U(f, P) = \sum_{k=1}^n M_k \Delta x_k.$$

Clearly, we have

$$L(f, P) \leq S(f, P, \xi) \leq U(f, P)$$

although the lower and upper sums need not be Riemann sums since m_k and M_k may not be in the range of f .

Let P and P_* be partitions of $[a, b]$. If $P \subseteq P_*$ then P_* is called a refinement of P . Thus, we can produce a refinement of P by inserting additional partition points between those in P . It is obvious that if P_* is a refinement of P then $\|P_*\| \leq \|P\|$. Also, if P and P^* are partitions of $[a, b]$, then they have a common refinement; for example, $P \cup P^*$ is a refinement of both P and P^* because it contains both sets of partition points.

Lemma 2.2.3. *Let P and P^* be partitions of $[a, b]$. Then $L(f, P) \leq U(f, P^*)$.*

Proof. Note that if partition points are added to a partition P then the lower sums increase and the upper sums decrease. In fact, if $L(f, P) = \sum_{k=1}^n m_k \Delta x_k$ and we add a point $z \in (x_{j-1}, x_j)$ to form $P_* = P \cup \{z\}$, then

$$L(f, P_*) = \sum_{k \neq j}^n m_k \Delta x_k + m_{\dagger}(z - x_{j-1}) + m_{\ddagger}(x_j - z),$$

where m_{\dagger} and m_{\ddagger} are the greatest lower bounds of f on $[x_{j-1}, z]$ and $[z, x_j]$ respectively. Thus $m_{\dagger} \geq m_j$ and $m_{\ddagger} \geq m_j$, so

$$m_{\dagger}(z - x_{j-1}) + m_{\ddagger}(x_j - z) \geq m_j(z - x_{j-1}) + m_j(x_j - z) = m_j(x_j - x_{j-1}).$$

Hence $L(f, P_*) \geq L(f, P)$. Similarly, $U(f, P_*) \leq U(f, P)$.

Suppose now that P and P^* are any two partitions of $[a, b]$ and $P_* = P \cup P^*$. Since P_* is a refinement of both P and P^* , the preceding observation gives

$$L(f, P) \leq L(f, P_*) \leq U(f, P_*) \leq U(f, P^*),$$

as desired. □

In the light of the foregoing lemma, we may define

$$\lambda(f) = \sup_P L(f, P) \quad \text{and} \quad \Lambda(f) = \inf_P U(f, P),$$

where both the supremum and the infimum are taken over all partitions P of $[a, b]$. From Lemma 2.2.3 it follows that $\lambda(f) \leq \Lambda(f)$. Moreover, the forthcoming lemma gives the necessary connection between the limit concept as $\|P\|$ tends to 0 and the supremum and infimum concepts.

Lemma 2.2.4. *If f is a bounded function on $[a, b]$, then*

$$\lim_{\|P\| \rightarrow 0} L(f, P) = \lambda(f) \quad \text{and} \quad \lim_{\|P\| \rightarrow 0} U(f, P) = \Lambda(f).$$

Proof. It suffices to prove that $\lim_{\|P\| \rightarrow 0} U(f, P) = \Lambda(f)$. Suppose that $|f(x)| < \kappa$ is valid for all $x \in [a, b]$ where $\kappa > 0$ is a constant. By the definition of $\Lambda(f)$, for any $\varepsilon > 0$ there exists a partition $P^* = \{z_l\}_{l=0}^j$ of $[a, b]$ such that

$$U(f, P^*) < \Lambda(f) + 2^{-1}\varepsilon.$$

Defining $\delta = \varepsilon/(4j\kappa)$, we show that if $\|P\| < \delta$ for any partition P of $[a, b]$ then $U(f, P) < \Lambda(f) + \varepsilon$. To do so, let $P_* = P \cup P^* = \{x_k\}_{k=0}^n$. Then some of the partition points of P are z_l 's from P^* , and some are not. Consequently,

$$U(f, P) = \sum_{[x_{k-1}, x_k] \cap P^* = \emptyset} M_k \Delta x_k + \sum_{[x_{k-1}, x_k] \cap P^* \neq \emptyset} M_k \Delta x_k.$$

Since the terms of the first sum on the right-hand side of the last representation are also terms of $U(f, P)$, it follows that $U(f, P) - U(f, P_*)$ consists of at most $2j$ terms (one for each of z_0 and z_j , and two for each of z_1, \dots, z_{j-1}), each of which does not exceed

$$\kappa \|P_*\| \leq \kappa \|P\| < \kappa \delta = (\kappa \varepsilon)(4j\kappa)^{-1} = \varepsilon(4j)^{-1}.$$

Therefore

$$U(f, P) - U(f, P_*) < 2j\varepsilon(4j)^{-1} = 2^{-1}\varepsilon.$$

Accordingly, we have

$$\Lambda(f) \leq U(f, P) < U(f, P_*) + 2^{-1}\varepsilon \leq U(f, P^*) + 2^{-1}\varepsilon < \Lambda(f) + \varepsilon,$$

whence completing the proof. \square

Because the Riemann sums lie between the upper and lower sums, it is natural to ask what happens if $\lambda(f) = \Lambda(f)$. One would expect that this forces the Riemann sums to converge, with their limit being equal to the common value of $\lambda(f)$ and $\Lambda(f)$. That is precisely what does happen, and it provides the following Darboux characterization of integrability.

Theorem 2.2.5. *Let $f : [a, b] \rightarrow \mathbf{R}$ be bounded. The following statements are equivalent.*

(i) $f \in R[a, b]$.

(ii) $\lambda(f) = \Lambda(f)$.

(iii) For any $\varepsilon > 0$ there is a $\delta > 0$ such that $U(f, P) - L(f, P) < \varepsilon$ as $\|P\| < \delta$.

Proof. (ii) \Rightarrow (iii) \Rightarrow (i). Suppose that $\lambda(f) = \Lambda(f) = s$ and $\varepsilon > 0$. Using Lemma 2.2.4, we can choose $\delta > 0$ so that

$$\|P\| < \delta \Rightarrow L(f, P) > s - 2^{-1}\varepsilon \text{ and } U(f, P) < s + 2^{-1}\varepsilon.$$

Then for any choice of $\{\xi_k\}_{k=0}^n$ in $[a, b]$, we have

$$s - 2^{-1}\varepsilon < L(f, P) \leq S(f, P, \xi) \leq U(f, P) < s + 2^{-1}\varepsilon.$$

Thus,

$$\begin{aligned} \|P\| < \delta &\Rightarrow U(f, P) - L(f, P) < \varepsilon \\ &\Rightarrow |S(f, P, \xi) - s| < \varepsilon \\ &\Rightarrow f \in R[a, b] \text{ and } \int_a^b f(x) dx = s. \end{aligned}$$

(i) \Rightarrow (iii) \Rightarrow (ii). If $f \in R[a, b]$ and $\int_a^b f(x) dx = s$, then for any $\varepsilon > 0$ we can find $\delta > 0$ such that

$$\|P\| < \delta \Rightarrow |S(f, P, \xi) - s| < 2^{-2}\varepsilon.$$

By the definitions of M_k and m_k , there exist points $\xi'_k, \xi''_k \in [x_{k-1}, x_k]$ satisfying

$$f(\xi'_k) > M_k - 2^{-2}\varepsilon(b-a)^{-1} \text{ and } f(\xi''_k) < m_k + 2^{-2}\varepsilon(b-a)^{-1}.$$

Then

$$\begin{aligned}
 U(f, P) &= \sum_{k=1}^n M_k \Delta x_k \\
 &< \sum_{k=1}^n \left(f(\xi'_k) + 2^{-2} \epsilon (b-a)^{-1} \right) \Delta x_k \\
 &= S(f, P, \hat{\xi}) + 2^{-2} \epsilon (b-a)^{-1} \sum_{k=1}^n \Delta x_k \\
 &< s + 2^{-1} \epsilon.
 \end{aligned}$$

Similarly, we have $L(f, P) > s - 2^{-1} \epsilon$, thereby getting

$$U(f, P) - L(f, P) < \epsilon \quad \text{and} \quad 0 \leq \Lambda(f) - \lambda(f) \leq U(f, P) - L(f, P) < \epsilon.$$

This, along with Lemma 2.2.4, implies $\lambda(f) = \Lambda(f)$. □

Remark 2.2.6. *It is worth remarking that $\int_a^b f(x) dx = \lambda(f) = \Lambda(f)$ holds for $f \in R[a, b]$.*

We have known that $R[a, b]$ is closed under addition and constant multiplication. To completely establish the algebraic closure of $R[a, b]$, we now use Theorem 2.2.5 to prove the result for the product of two general functions in $R[a, b]$.

Theorem 2.2.7. *If $f_1, f_2 \in R[a, b]$, then $f_1 f_2 \in R[a, b]$.*

Proof. For convenience, let $f = f_1$ and $g = f_2$. We first prove the assertion for the special case: $f, g \geq 0$. Suppose that P is any partition of $[a, b]$, and let $M_{f,k}$, $M_{g,k}$ and $M_{fg,k}$ denote the least upper bounds of f , g , and fg , respectively, on the k -th subinterval $[x_{k-1}, x_k]$. It is not hard to see that $M_{fg,k} \leq M_{f,k} M_{g,k}$. Similarly, if $m_{f,k}$, $m_{g,k}$, and $m_{fg,k}$ are the corresponding greatest lower bounds on $[x_{k-1}, x_k]$, then $m_{fg,k} \geq m_{f,k} m_{g,k}$. Thus

$$M_{fg,k} - m_{fg,k} \leq M_{f,k} M_{g,k} - m_{f,k} m_{g,k}.$$

Now, let

$$M_f = \sup_{x \in [a, b]} f(x) \quad \text{and} \quad M_g = \sup_{x \in [a, b]} g(x),$$

and rewrite the last inequality as

$$\begin{aligned} M_{fg,k} - m_{fg,k} &\leq M_{f,k}M_{g,k} - m_{f,k}M_{g,k} + m_{f,k}M_{g,k} - m_{f,k}m_{g,k} \\ &\leq (M_{f,k} - m_{f,k})M_{g,k} + (M_{g,k} - m_{g,k})M_{f,k}. \end{aligned}$$

Accordingly, we have

$$\begin{aligned} U(fg, P) - L(fg, P) &= \sum_{k=1}^n (M_{fg,k} - m_{fg,k}) \Delta x_k \\ &\leq M_g \sum_{k=1}^n (M_{f,k} - m_{f,k}) \Delta x_k + M_f \sum_{k=1}^n (M_{g,k} - m_{g,k}) \Delta x_k \\ &= M_g (U(f, P) - L(f, P)) + M_f (U(g, P) - L(g, P)). \end{aligned}$$

By Theorem 2.2.5, we have that if $\varepsilon > 0$ then $f, g \in R[a, b]$ allows us to choose P so that

$$U(f, P) - L(f, P) < 2^{-1}\varepsilon(1 + M_g)^{-1} \text{ and } U(g, P) - L(g, P) < 2^{-1}\varepsilon(1 + M_f)^{-1}.$$

This gives

$$U(fg, P) - L(fg, P) < \varepsilon \text{ and so } fg \in R[a, b].$$

To prove the general case in which f and g need not be nonnegative, we use Theorem 2.1.8 which assures us that f and g are bounded from below, say, by constants κ_f and κ_g respectively. Then $f - \kappa_f, g - \kappa_g \geq 0$ on $[a, b]$ and $f - \kappa_f, g - \kappa_g \in R[a, b]$, so the just-proved case can be applied to conclude that $(f - \kappa_f)(g - \kappa_g) \in R[a, b]$. Note that

$$fg = (f - \kappa_f)(g - \kappa_g) + \kappa_f g + \kappa_g f + \kappa_f \kappa_g$$

for which each term of the right-hand member is in $R[a, b]$. Thus it follows from the above special case, as well as Theorem 2.2.1, that $fg \in R[a, b]$. \square

It may be observed that no explicit formula is valid for $\int_a^b f_1(x)f_2(x)dx$. But, there is an inequality to connect it with $\int_a^b f_1^2(x)dx$ and $\int_a^b f_2^2(x)dx$. This is just the Cauchy-Schwarz inequality or the Cauchy-Bunyakovski-Schwarz inequality, named after Augustin Louis Cauchy, Hermann Amandus Schwarz and Viktor Yakovlevich Bunyakovsky.

Theorem 2.2.8. *If $f_1, f_2 \in R[a, b]$, then*

$$\left| \int_a^b f_1(x)f_2(x) dx \right| \leq \left(\int_a^b f_1^2(x) dx \right)^{2^{-1}} \left(\int_a^b f_2^2(x) dx \right)^{2^{-1}}.$$

Proof. It is clear that

$$f_1f_2, f_1^2, f_2^2 \in R[a, b] \text{ and } 0 \leq \int_a^b f_1^2(x) dx, \int_a^b f_2^2(x) dx.$$

If

$$\begin{cases} \alpha = \int_a^b f_1^2(x) dx; \\ \beta = 2 \int_a^b f_1(x)f_2(x) dx; \\ \gamma = \int_a^b f_2^2(x) dx, \end{cases}$$

then

$$\tau(t) = \alpha t^2 + \beta t + \gamma = \int_a^b (tf_1(x) + f_2(x))^2 dx \geq 0.$$

From the elementary algebra we recall that $\tau(t) = 0$ has two roots in \mathbf{R} if and only if $\beta^2 - 4\alpha\gamma > 0$. Since $Q(t)$ is never negative, the quadratic equation $\tau(t) = 0$ can not have two real roots, which means that $\beta^2 - 4\alpha\gamma \leq 0$. But this says that

$$\left(\int_a^b f_1(x)f_2(x) dx \right)^2 - \left(\int_a^b f_1^2(x) dx \right) \left(\int_a^b f_2^2(x) dx \right) \leq 0.$$

□

2.3 Fundamental Theorem of Calculus

This section is devoted to proving the fundamental theorem of calculus via an investigation of the integrability of continuous functions.

The ordinary definition of continuity is actually pointwise continuity and allows us to choose δ (in terms of the $\varepsilon - \delta$ language) in a point-by-point manner, using different choices of δ for different points in the domain of f . But sometimes, say, $f(x) = x^2$ on $[-5, 5]$, or $f(x) = x^{-1}$ on $[1, \infty)$, we are able to choose one δ that works uniformly well at any point in the domain of f . This provides both the motivation and the name of the next concept.

Definition 2.3.1. *Given a set $D \subseteq \mathbf{R}$ and a function $f : D \rightarrow \mathbf{R}$, we say that:*

(i) f has a limit $L \in \mathbf{R}$ at $c \in D$, denoted $\lim_{x \rightarrow c} f(x) = L$, provided that for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$x \in D \text{ and } 0 < |x - c| < \delta \Rightarrow |f(x) - L| < \varepsilon;$$

(ii) f is continuous on D provided that it is continuous at any point $c \in D$ -i.e.- for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$x \in D \text{ and } |x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon;$$

(iii) f is uniformly continuous on D provided that for any $\varepsilon > 0$ there exists a $\delta > 0$ (depending only on ε and D) such that

$$x_1, x_2 \in D \text{ and } |x_1 - x_2| < \delta \Rightarrow |f(x_1) - f(x_2)| < \varepsilon.$$

It is plain that the linear function $x \mapsto f(x) = ax + b$ with $a \neq 0$ is uniformly continuous on \mathbf{R} : for $\varepsilon > 0$, take $\delta = \varepsilon|a|^{-1}$. However, the sign function $f(x) = |x|x^{-1}$ is not uniformly continuous on $\mathbf{R} \setminus \{0\}$ although it is so on $(-\infty, 0)$ and $(0, \infty)$. Also, it is easy to see that if f is uniformly continuous on D , then f is continuous on D . Though its converse is not true in general, there are situations in which continuity at every point of D does imply uniform continuity on D .

Unless a special remark is made, the symbol $C[a, b]$ is employed to represent the set of all real-valued continuous functions on $[a, b]$.

Theorem 2.3.2. *Below are the basic properties of $f \in C[a, b]$.*

- (i) f is bounded; that is, $\sup\{|f(x)| : x \in [a, b]\} < \infty$.
- (ii) f has a minimum and a maximum; that is, there exist $x_1, x_2 \in [a, b]$ such that $f(x_1) = \sup\{f(x) : x \in [a, b]\}$ and $f(x_2) = \inf\{f(x) : x \in [a, b]\}$.
- (iii) *Intermediate value theorem* – for any y between $f(a)$ and $f(b)$ there is $x \in [a, b]$ such that $f(x) = y$.
- (iv) f is uniformly continuous on $[a, b]$.

Proof. (i) If not, then for any $n \in \mathbf{N}$ there is an $x_n \in [a, b]$ such that $|f(x_n)| \geq n$. By the Bolzano-Weierstrass property – see also Theorem 1.2.4 (ii), there exist $x \in [a, b]$ and a subsequence $\{x_{n_k}\}_{k \in \mathbf{N}}$ of $\{x_n\}_{n \in \mathbf{N}}$ such that $x_{n_k} \rightarrow x$. Hence, $f \in C[a, b]$ yields $\lim_{k \rightarrow \infty} f(x_{n_k}) = f(x)$ and so $\{f(x_{n_k})\}_{k \in \mathbf{N}}$ is bounded. This is a contradiction since $|f(x_{n_k})| \geq n_k \geq k$ for any $k \in \mathbf{N}$. Therefore, f must be bounded on $[a, b]$.

(ii) By (i), $M = \sup\{f(x) : x \in [a, b]\}$ exists in \mathbf{R} . This produces a sequence $x_n \in [a, b]$ with $M - n^{-1} < f(x_n) \leq M$ for any $n \in \mathbf{N}$. Again by the Bolzano-Weierstrass property, there is a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ of $\{x_n\}_{n=1}^{\infty}$ such that $x_{n_k} \rightarrow x \in [a, b]$. Accordingly, from $f \in C[a, b]$ and Theorem 1.2.7 it follows that

$$\begin{cases} f(x_{n_k}) \rightarrow f(x); \\ f(x_{n_k}) \rightarrow M; \\ M = f(x). \end{cases}$$

The argument for a minimum is similar.

(iii) It suffices to consider the case $f(a) \neq f(b)$. Of course, we may assume $f(a) < y < f(b)$ and $S = \{x' \in [a, b] : f(x') < y\}$. Since $a \in S$, it follows that $S \neq \emptyset$ and it has an upper bound b . The axiom of LUB for \mathbf{R} implies that $x = \sup S$ exists. The definition of S gives $a < x < b$. Next, we prove $f(x) = y$ through handling two cases below.

Case 1: $f(x) < y$. Since $f \in C[a, b]$, there is a $\delta > 0$ such that

$$|x_1 - x| < \delta \text{ and } x_1 \in [a, b] \Rightarrow f(x_1) < y.$$

Consequently, there is an $x_2 \in [a, b]$ with $x_2 > x$ and $f(x_2) < y$, contradicting the definition of x .

Case 2: $f(x) > y$. Like *Case 1*, we can find a $\delta > 0$ such that

$$x - \delta < x_3 < x \text{ and } x_3 \in [a, b] \Rightarrow f(x_3) > y,$$

which is against the choice of x .

(iv) Let $c \in [a, b]$. Since $f \in C[a, b]$, for any $\varepsilon > 0$ there is a $\delta_c > 0$ such that

$$|x - c| < \delta_c \Rightarrow |f(x) - f(c)| < 2^{-1}\varepsilon.$$

Now consider the collection of open intervals:

$$O = \{(c - 2^{-1}\delta_c, c + 2^{-1}\delta_c) : c \in [a, b]\}.$$

For each $c \in [a, b]$, there is an interval in O centered at c , so O is obviously an open cover of $[a, b]$. By the Heine-Borel property – see also Theorem 1.2.4 (iii), there exists a finite subcollection of O that covers $[a, b]$, say,

$$[a, b] \subset \bigcup_{j=1}^n (c_j - 2^{-1}\delta_{c_j}, c_j + 2^{-1}\delta_{c_j}).$$

Define $\delta = 2^{-1} \min_{1 \leq j \leq n} \delta_{c_j}$. Now if $|x_1 - x_2| < \delta$ for $x_1, x_2 \in [a, b]$, then x_1 is in one of the n intervals in the above finite covering, so $|x_1 - c_j| < \delta$ holds for some c_j . Consequently,

$$|x_2 - c_j| \leq |x_2 - x_1| + |x_1 - c_j| < \delta + 2^{-1}\delta_{c_j} \leq \delta_{c_j}.$$

Thus, the choice of δ_{c_j} yields

$$|f(x_1) - f(c_j)| < 2^{-1}\epsilon \quad \text{and} \quad |f(x_2) - f(c_j)| < 2^{-1}\epsilon,$$

and consequently,

$$|f(x_1) - f(x_2)| \leq |f(x_1) - f(c_j)| + |f(x_2) - f(c_j)| < \epsilon.$$

□

Example 2.3.3. To handle the integrability of continuous functions, we consider the integrability of the following function

$$f(x) = \begin{cases} q^{-1}, & x = mn^{-1} \in \mathbf{Q} \cap (0, 1] \text{ and } mn^{-1} \text{ in lowest terms} \\ 0, & x \in \{0\} \cup ((0, 1] \cap (\mathbf{R} \setminus \mathbf{Q})) \end{cases}.$$

Recall that this function f is continuous at each irrational number and discontinuous at each rational number in $(0, 1)$: in fact, if $c \in (0, 1)$ is rational, then we take irrational points $s_k \in (0, 1)$ converging to c and hence $\lim_{k \rightarrow \infty} f(s_k) = 0 \neq f(c)$; in order to see that f is continuous at any irrational $d \in (0, 1)$, let $\epsilon > 0$ and note that there are only a finite number of irreducible fractions mn^{-1} in $(0, 1)$ such that $n^{-1} \geq \epsilon$. Therefore there is a $\delta > 0$ such that those fractions do not belong to $(d - \delta, d + \delta)$. This means that for all x satisfying $|x - d| < \delta$, we have $0 \leq f(x) < \epsilon$. It is worth mentioning that f is continuous from the right at 0 and discontinuous from the left at 1.

We now assert that $f \in R[0, 1]$ with $\int_0^1 f(x) dx = 0$. As a matter of fact, the density of $\mathbf{R} \setminus \mathbf{Q}$ in \mathbf{R} ensures that $L(f, P) = 0$ for every partition P of $[0, 1]$. Thus, for any $\epsilon > 0$, we seek a partition P such that $U(f, P) < \epsilon$. In $[0, 1]$ there are just finitely many points p/q such that $n^{-1} > 2^{-1}\epsilon$. Let l be the number of such points in $[0, 1]$ and let P be a partition such that $\|P\| < \epsilon(4l)^{-1}$. In $U(f, P) = \sum_{k=1}^n M_k \Delta x_k$ there are at most $2l$ terms where $M_k > 2^{-1}\epsilon$; for each of these terms we have $M_k \Delta x_k \leq \|P\| < \epsilon(4l)^{-1}$. Therefore the total of these terms is less than $(2l)(\epsilon(4l)^{-1}) = 2^{-1}\epsilon$. In each of the remaining terms, $M_k \leq 2^{-1}\epsilon$, so their total is less than $(2^{-1}\epsilon) \sum_{k=1}^n \Delta x_k = 2^{-1}\epsilon$. Hence $U(f, P) < \epsilon$, so the Darboux's integrability criterion implies that $f \in R[0, 1]$ and $\int_0^1 f(x) dx = 0$.

Up to this point, we have seen very few examples that we could verify as being integrable functions. But, the next two theorems provide us with a very large set of such examples.

Theorem 2.3.4. *If $f \in C[a, b]$, then $f \in R[a, b]$.*

Proof. By Theorem 2.3.2 (iv), f is uniformly continuous on $[a, b]$, and so for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$x_{\dagger}, x_{\ddagger} \in [a, b] \text{ and } |x_{\dagger} - x_{\ddagger}| < \delta \Rightarrow |f(x_{\dagger}) - f(x_{\ddagger})| < \varepsilon(b-a)^{-1}.$$

Now choose a partition P such that $\|P\| < \delta$, and consider the k -th subinterval $[x_{k-1}, x_k]$. There are $\xi_{k,\dagger}, \xi_{k,\ddagger} \in [x_{k-1}, x_k]$ such that $f(\xi_{k,\dagger}) = M_k, f(\xi_{k,\ddagger}) = m_k$ thanks to Theorem 2.3.2 (ii). Note that $|\xi_{k,\dagger} - \xi_{k,\ddagger}| \leq \Delta x_k \leq \|P\| < \delta$. So it follows that $M_k - m_k < \varepsilon(b-a)^{-1}$. Thus

$$U(f, P) - L(f, P) = \sum_{k=1}^n (M_k - m_k) \Delta x_k < \varepsilon;$$

and hence $f \in R[a, b]$ by Theorem 2.2.5. □

As a consequence of Theorem 2.3.4, we can establish the mean value theorem for integrals.

Theorem 2.3.5. *If $f \in C[a, b]$, then there is a number $\xi \in [a, b]$ such that $f(\xi)(b-a) = \int_a^b f(x) dx$.*

Proof. By Theorem 2.3.4, we have $f \in R[a, b]$. Let

$$m = \min_{x \in [a, b]} f(x) \text{ and } M = \max_{x \in [a, b]} f(x).$$

Then

$$m(b-a) \leq \int_a^b f(x) dx \leq M(b-a);$$

that is, $(b-a)^{-1} \int_a^b f(x) dx$ lies between m and M . Since $f \in C[a, b]$, f assumes any value between m and M and so there is a $\xi \in [a, b]$ such that

$$f(\xi) = (b-a)^{-1} \int_a^b f(x) dx.$$

□

The idea involved in the argument of Theorem 2.3.5 also leads to the following result.

Theorem 2.3.6. *If f is monotonic on $[a, b]$; that is, f is increasing*

$$x_1 \leq x_2 \text{ and } x_1, x_2 \in [a, b] \Rightarrow f(x_1) \leq f(x_2)$$

or decreasing

$$x_1 \leq x_2 \text{ and } x_1, x_2 \in [a, b] \Rightarrow f(x_1) \geq f(x_2),$$

then $f \in R[a, b]$.

Proof. Without loss of generality, we may assume that f is increasing and non-constant (otherwise, there will be nothing to prove). Then for any subinterval $[x_{k-1}, x_k]$ of $[a, b]$, M_k occurs at the right endpoint: $M_k = f(x_k)$. Similarly, $m_k = f(x_{k-1})$. Therefore, if P is any partition of $[a, b]$, then

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{k=1}^n (M_k - m_k) \Delta x_k \\ &\leq \|P\| \sum_{k=1}^n (f(x_k) - f(x_{k-1})) \\ &= \|P\| |f(b) - f(a)|. \end{aligned}$$

Now if $\varepsilon > 0$, then we simply select a partition P such that

$$\|P\| < \varepsilon (f(b) - f(a) + 1)^{-1} \text{ and hence } U(f, P) - L(f, P) < \varepsilon.$$

This infers that $f \in R[a, b]$ by Theorem 2.2.5. □

To get the fundamental theorem of calculus linking an integral to a derivative, we recall the notion of differentiation and the corresponding mean value theorem.

Definition 2.3.7. *For an interval $I \subseteq \mathbf{R}$ containing an element c , we say that the function $f : I \rightarrow \mathbf{R}$ is differentiable at c or has a derivative at c provided that the limit*

$$f'(c) = \lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c}$$

exists; that is, for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$x, c \in I \text{ and } 0 < |x - c| < \delta \Rightarrow \left| f'(c) - \frac{f(x) - f(c)}{x - c} \right| < \varepsilon.$$

If f is differentiable at every element of the set $D \subseteq I$, then f is called differentiable on D , and the function $f' : D \rightarrow \mathbf{R}$ is called the derivative of f on D .

Obviously, we can read off that if f is differentiable at c then it is continuous but not conversely. At the same time, we can take the derivative of the sum, product, quotient and composition of two differentiable functions. Moreover we can establish the generalized mean value theorem for derivatives as follows.

Theorem 2.3.8. *Let $f, g \in C[a, b]$ be differentiable on (a, b) . Then there exists a point $c \in (a, b)$ such that $(f(b) - f(a))g'(c) = (g(b) - g(a))f'(c)$.*

Proof. Define the function $h : [a, b] \rightarrow \mathbf{R}$ by

$$h(x) = (f(b) - f(a))g(x) - (g(b) - g(a))f(x).$$

Then $h \in C[a, b]$ and is differentiable on (a, b) with $h(a) = h(b)$. An application of Theorem 2.3.2 (ii) produces two points $x_1, x_2 \in [a, b]$ such that

$$\inf\{h(x) : x \in [a, b]\} = h(x_1) \text{ and } \sup\{h(x) : x \in [a, b]\} = h(x_2).$$

If x_1 and x_2 are both endpoints of $[a, b]$, then $h(x) = h(a) = h(b)$ for any $x \in [a, b]$, and hence f is a constant, giving $h'(x) = 0$ for all $x \in (a, b)$. Otherwise, we may assume $x_2 \in (a, b)$. Now let $s_n = x_2 - n^{-1}$ be a sequence converging to x_2 such that $a < s_n < x_2$ for sufficiently large $n \in \mathbf{N}$. Since h is differentiable at x_2 , we have

$$0 \leq \frac{h(s_n) - h(x_2)}{s_n - x_2} \rightarrow h'(x_2) \text{ as } n \rightarrow \infty,$$

whence $h'(x_2) \geq 0$. Similarly, if $x_2 < t_n = x_2 + n^{-1} < b$, then

$$0 \geq \frac{h(t_n) - h(x_2)}{t_n - x_2} \rightarrow h'(x_2) \text{ as } n \rightarrow \infty,$$

and hence $h'(x_2) \leq 0$. Therefore, we have $h'(x_2) = 0$, whence

$$(f(b) - f(a))g'(x_2) = (g(b) - g(a))f'(x_2).$$

□

The coming-up-next is the fundamental theorem of calculus whose importance can hardly be overstated.

Theorem 2.3.9. *If $f \in R[a, b]$ and $F'(x) = f(x)$ for $x \in [a, b]$, then $\int_a^b f(x) dx = F(b) - F(a)$. Conversely, if $f \in C[a, b]$ and $F(x) = \int_a^x f(t) dt$ for $x \in [a, b]$, then $F'(x) = f(x)$ for $x \in [a, b]$.*

Proof. Let P be a partition of $[a, b]$, and consider the collapsing sum

$$F(b) - F(a) = \sum_{k=1}^n (F(x_k) - F(x_{k-1})).$$

Since F is differentiable on $[x_{k-1}, x_k]$, we can apply the special case (where $g(x) = x$) of Theorem 2.3.8 to get a number $\xi_k \in [x_{k-1}, x_k]$ such that

$$F(x_k) - F(x_{k-1}) = F'(\xi_k) \Delta x_k = f(\xi_k) \Delta x_k.$$

Via substituting the right-hand member of this equation into the collapsing sum, we get $F(b) - F(a) = S(f, P, \xi)$, a Riemann sum for f with respect to P . Thus we have shown that for any partition P , the points $\{\xi_k\}_{k=1}^n$ can be chosen so that $S(f, P, \xi)$ is $F(b) - F(a)$. Therefore $F(b) - F(a)$ is the only possible value of $\lim_{\|P\| \rightarrow 0} S(f, P, \xi)$. Since $f \in R[a, b]$, this limit must exist and its value is denoted by $\int_a^b f(x) dx$. Thus $\int_a^b f(x) dx = F(b) - F(a)$.

To prove the converse, let $c \in [a, b]$ and consider the difference quotient

$$Q_c(h) = h^{-1} (F(c+h) - F(c)) = h^{-1} \int_c^{c+h} f(t) dt.$$

By Theorem 2.3.5, there is a ξ between c and $c+h$ such that

$$\int_c^{c+h} f(t) dt = f(\xi)h.$$

Via substituting this into the expression for $Q_c(h)$, we obtain $Q_c(h) = f(\xi)$. As $h \rightarrow 0$, it follows that $\xi \rightarrow c$; therefore

$$F'(c) = \lim_{h \rightarrow 0} Q_c(h) = \lim_{\xi \rightarrow c} f(\xi) = f(c),$$

thanks to the continuity of f at c . □

Theorem 2.3.9 is a powerful tool for the computation of integrals. For example, if $n \in \mathbf{N}$, then $(n+1)^{-1}x^{n+1}$ has the continuous derivative x^n , and hence

$$\int_a^b x^n dx = (n+1)^{-1}(b^{n+1} - a^{n+1}).$$

Additionally, we should notice that the condition $F' = f$ in the first part of Theorem 2.3.9 cannot guarantee $f \in R[a, b]$; for instance, if

$$F(x) = \begin{cases} x^2 \sin x^{-2}, & x \neq 0 \\ 0, & x = 0 \end{cases},$$

then

$$F'(x) = \begin{cases} 2x \sin x^{-2} - 2x^{-1} \cos x^{-2}, & x \neq 0 \\ 0, & x = 0 \end{cases},$$

and hence $f = F' \notin R[-1, 1]$ due to the fact that f is unbounded on $[-1, 1]$.

As a consequence of Theorem 2.3.9, we can establish the following formula for change of variable for the Riemann integrals.

Corollary 2.3.10. *Let U and V be open intervals in \mathbf{R} , $\phi : U \rightarrow V$ a differentiable function with continuous derivative, and $f : V \rightarrow \mathbf{R}$ a continuous function. Then*

$$\int_{\phi(a)}^{\phi(b)} f(v) dv = \int_a^b f(\phi(u))\phi'(u) du \text{ for all } a, b \in U.$$

Proof. Let $F : V \rightarrow \mathbf{R}$ be the function defined by $F(y) = \int_{\phi(a)}^y f(v) dv$ for all $y \in V$. Then F is differentiable and $F' = f$ by Theorem 2.3.9. The function $G : U \rightarrow \mathbf{R}$ defined by $G(x) = \int_{\phi(a)}^{\phi(x)} f(v) dv$ is actually the composition $G = F \circ \phi$ of two differentiable functions, hence is itself differentiable. By the chain rule we have

$$G'(x) = F'(\phi(x))\phi'(x) = f(\phi(x))\phi'(x) \text{ for all } x \in U,$$

thereby getting

$$G(x) = \int_a^x f(\phi(u))\phi'(u) du + C \text{ for some constant } C \in \mathbf{R}.$$

Upon setting $x = a$, we achieve

$$C = 0 \text{ and } G(x) = \int_a^x f(\phi(u))\phi'(u) du \text{ for all } x \in U.$$

Letting $x = b$ gives the desired formula. □

2.4 Improper Integrals

Since there are many useful, commonly encountered functions whose domains are unbounded or whose graphs have vertical asymptotes, in this section we extend the Riemann integrals to cover two new types of integrals for some of these functions. As one of their applications, we prove the Weierstrass approximation theorem via the convolution of two functions on \mathbf{R} .

Definition 2.4.1.

(i) Let $a \in \mathbf{R}$ and $f \in R[a, t]$ for every $t \geq a$. Then the improper integral of f on $[a, \infty)$ is defined by

$$\int_a^\infty f(x) dx = \lim_{t \rightarrow \infty} \int_a^t f(x) dx.$$

If the limit exists; that is, for any $\varepsilon > 0$ there is a $t_0 > 0$ such that

$$t > t_0 \Rightarrow \left| \int_a^t f(x) dx - \int_a^\infty f(x) dx \right| < \varepsilon,$$

then the improper integral is said to be convergent; otherwise, the improper integral is divergent.

(ii) The improper integral of a function f on $(-\infty, b]$ is defined similarly:

$$\int_{-\infty}^b f(x) dx = \lim_{t \rightarrow -\infty} \int_t^b f(x) dx.$$

(iii) If $f \in R[-t, t]$ for any $t > 0$, then the improper integral of f on $(-\infty, \infty)$ is defined by

$$\int_{-\infty}^\infty f(x) dx = \lim_{t \rightarrow -\infty} \int_t^c f(x) dx + \lim_{t \rightarrow \infty} \int_c^t f(x) dx \text{ for all } c \in \mathbf{R}.$$

If $\int_{-\infty}^c f(x) dx$ and $\int_c^\infty f(x) dx$ are convergent, then

$$\int_{-\infty}^\infty f(x) dx = \int_{-\infty}^c f(x) dx + \int_c^\infty f(x) dx.$$

The following examples show how to evaluate these improper integrals.

Example 2.4.2.

(i)

$$\int_1^{\infty} x^{-2} dx = \lim_{t \rightarrow \infty} \int_1^t x^{-2} dx = \lim_{t \rightarrow \infty} (1 - t^{-1}) = 1.$$

Hence the improper integral is convergent.

(ii)

$$\int_{-\infty}^0 \cos x dx = \lim_{t \rightarrow -\infty} \int_t^0 \cos x dx = \lim_{t \rightarrow -\infty} (-\sin t).$$

Hence the improper integral is divergent.

(iii)

$$\begin{aligned} \int_{-\infty}^{\infty} (1+x^2)^{-1} dx &= \int_{-\infty}^0 (1+x^2)^{-1} dx + \int_0^{\infty} (1+x^2)^{-1} dx \\ &= \lim_{t \rightarrow -\infty} \arctan x \Big|_t^0 + \lim_{t \rightarrow \infty} \arctan x \Big|_0^t \\ &= \pi. \end{aligned}$$

Hence the improper integral is convergent.

(iv)

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-x} dx &= \int_{-\infty}^0 e^{-x} dx + \int_0^{\infty} e^{-x} dx \\ &= \lim_{t \rightarrow -\infty} (-e^{-x}) \Big|_t^0 + \lim_{t \rightarrow \infty} (-e^{-x}) \Big|_0^t \\ &= \infty. \end{aligned}$$

Hence the improper integral is divergent.

Here is a fundamental result on the improper integrals of which (iii) is called Cauchy's convergence criterion.

Theorem 2.4.3. Let $\int_a^{\infty} f_1(x) dx$ and $\int_a^{\infty} f_2(x) dx$ exist. Then

(i) $\int_a^{\infty} (c_1 f_1(x) + c_2 f_2(x)) dx = c_1 \int_a^{\infty} f_1(x) dx + c_2 \int_a^{\infty} f_2(x) dx$ for any constants $c_1, c_2 \in \mathbf{R}$.

(ii) $\int_a^{\infty} f_1(x) dx \leq \int_a^{\infty} f_2(x) dx$ provided that $f_1 \leq f_2$ on $[a, \infty)$.

(iii) $\int_a^{\infty} f(x) dx$ exists if and only if for any $\varepsilon > 0$ there is a $t_0 > 0$ such that $b_2 > b_1 > t_0$ implies $\left| \int_{b_1}^{b_2} f(x) dx \right| < \varepsilon$.

Proof. As an exercise, the arguments for (i) and (ii) are left for the students. Now we verify (iii). If $\int_a^\infty f(x) dx$ exists, then for any $\varepsilon > 0$ there is a $t_0 > 0$ such that

$$t > t_0 \Rightarrow \left| \int_a^t f(x) dx - \int_a^\infty f(x) dx \right| < 2^{-1}\varepsilon.$$

Thus the desired inequality follows from taking $b_2 > b_1 > t_0$.

On the other hand, if for any $\varepsilon > 0$ there is a $t_0 > 0$ such that $b_2 > b_1 > t_0$ implies $\left| \int_{b_1}^{b_2} f(x) dx \right| < \varepsilon$, then the sequence $\left\{ \int_a^{a+n} f(x) dx \right\}_{n=1}^\infty$ is a Cauchy sequence, and hence converges to a finite number s . This gives that for $\varepsilon > 0$ there exists a $t_0 > 0$ such that

$$a + n \geq t_0 \Rightarrow \left| \int_a^{a+n} f(x) dx - s \right| < \varepsilon.$$

Then if $\mathbf{N} \ni n_0 > t_0 - a$, then $t > a + n_0$ yields

$$\left| \int_a^t f(x) dx - s \right| = \left| \int_a^{a+n_0} f(x) dx - s + \int_{a+n_0}^t f(x) dx \right| < 2\varepsilon,$$

as required. □

Defining the improper integral in terms of the definite Riemann integral has both advantages and disadvantages. On the positive side, there is no need to develop a separate limit theory for the new type of integral. But, on the other hand, we are required to evaluate $\int_a^t f(x) dx$ as a function of t in order to determine whether $\int_a^\infty f(x) dx$ is convergent, and the difficulty of finding the primitive of a given function is well-known to the students. Since it is frequently sufficient to determine convergence of $\int_a^\infty f(x) dx$ without finding its value, it will be very helpful to have a test that says that under some appropriate conditions the convergence of a simple integral implies that of a complicated one. Such a result is called a comparison test. That is the nature of the next theorem.

Theorem 2.4.4. *Let $f_1, f_2 \in R[a, t]$ and $0 \leq f_1(t) \leq f_2(t)$ for every $a \leq t < \infty$. If $\int_a^\infty f_2(x) dx$ is convergent, then $\int_a^\infty f_1(x) dx$ is convergent too.*

Proof. Let $f = f_1$ and $g = f_2$. First, we verify that if there is a number $M > 0$ such that $F(t) = \int_a^t f(x) dx \leq M$ for every $t \geq a$, then $\int_a^\infty f(x) dx$ is convergent. In fact, since

$$F(t+h) - F(t) = \int_t^{t+h} f(x) dx,$$

it follows that F is increasing on $[a, \infty)$. Therefore $F(t) \leq$ ensures that

$$\lim_{t \rightarrow \infty} F(t) = \sup_{t \geq a} F(t) < \infty, \text{ i.e., } \lim_{t \rightarrow \infty} \int_a^t f(x) dx \text{ exists.}$$

Secondly, if

$$h(x) = g(x) - f(x) \text{ for all } x \geq a,$$

then $h \in R[a, t]$ and $h(t) \geq 0$ for every $t \geq a$, and hence

$$\int_a^t h(x) dx \leq \int_a^t g(x) dx \leq \int_a^\infty g(x) dx.$$

Thus $\int_a^\infty g(x) dx$ is an upper bound for $\int_a^t h(x) dx$ for $t \geq a$, so by the first paragraph, $\int_a^\infty h(x) dx$ is convergent. Now

$$\lim_{t \rightarrow \infty} \int_a^t f(x) dx = \lim_{t \rightarrow \infty} \int_a^t g(x) dx - \lim_{t \rightarrow \infty} \int_a^t h(x) dx,$$

and the last two limits exist because both $\int_a^\infty g(x) dx$ and $\int_a^\infty h(x) dx$ are convergent. \square

The utility of this result is illustrated in the convergence of the improper integral

$$\int_1^\infty (1+x^3)^{-1} dx \text{ since } (1+x^3)^{-1} \leq x^{-2} \text{ for } x \geq 1.$$

Regarding the integrals of unbounded functions, we have the following definition.

Definition 2.4.5.

(i) Let $f \in R[c, d]$ for every $[c, d] \subset [a, b)$. Then the improper integral of f on $[a, b]$ is defined by

$$\int_a^{b-} f(x) dx = \lim_{t \rightarrow b-} \int_a^t f(x) dx.$$

If the limit exists; that is, for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$0 < b - t < \delta \Rightarrow \left| \int_a^{b-} f(x) dx - \int_a^t f(x) dx \right| < \varepsilon,$$

then the improper integral is said to be convergent; otherwise, the improper integral is divergent.

(ii) Similarly, we define the analogous improper integral:

$$\int_{a^+}^b f(x) dx = \lim_{t \rightarrow a^+} \int_t^b f(x) dx.$$

(iii) If f is unbounded at a and b , then for any $c \in (a, b)$, the improper integral $\int_{a^+}^{b^-} f(x) dx$ is defined by

$$\lim_{t \rightarrow a^+} \int_t^c f(x) dx + \lim_{t \rightarrow b^-} \int_c^t f(x) dx.$$

It is convergent if both $\int_{a^+}^c f(x) dx$ and $\int_c^{b^-} f(x) dx$ are convergent, and in that case

$$\int_{a^+}^{b^-} f(x) dx = \int_{a^+}^c f(x) dx + \int_c^{b^-} f(x) dx.$$

Example 2.4.6.

$$\int_0^1 (x^2 - 1)^{-1} dx \quad \text{and} \quad \int_{-1}^0 (x^2 - 1)^{-1} dx \quad \text{and} \quad \int_{-1}^1 (x^2 - 1)^{-1} dx,$$

are improper integrals satisfying the foregoing definition.

Plainly, the preceding types of improper integrals can be used in combination to form the improper integral of any function with a graph that has at most a finite number of asymptotes. For instance, if

$$f(x) = ((x-a)(x-b))^{-1},$$

where $a < b$, then we choose an arbitrary number $c \in (a, b)$ and write

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_{-\infty}^{a-1} f(x) dx + \int_{a-1}^{a^-} f(x) dx + \int_{a^+}^c f(x) dx \\ &\quad + \int_c^{b^-} f(x) dx + \int_{b^+}^{b+1} f(x) dx + \int_{b+1}^{\infty} f(x) dx. \end{aligned}$$

Though the calculation of such an integral may be cumbersome – we have just decomposed one improper integral into six limits – the major achievement here is that we do not need a separate theory for each situation. There are just two basic types of improper integrals, and finite combinations of these two types suffice for most functions.

Naturally, there is a comparison test for the improper integrals of unbounded functions. It is analogous to Theorem 2.4.4 and its proof can be achieved similarly.

Theorem 2.4.7. Let $f_1, f_2 \in R[a, t]$ and $f_1(t) \leq f_2(t)$ for every $a \leq t < b$. If $\int_a^{b-} f_2(x) dx$ is convergent, then $\int_a^{b-} f_1(x) dx$ is convergent too.

This theorem has an obvious dual for $\int_{a+}^b f(x) dx$, which is illustrated by

$$\int_{0+}^1 (x(x+1))^{-1} dx = \infty,$$

thanks to

$$(x(x+1))^{-1} \geq (2x)^{-1} \text{ for all } x \in (0, 1].$$

As a good application of the improper integrals, we consider the Gamma function.

Example 2.4.8. For each $x > 0$, let

$$\Gamma(x) = \int_{0+}^{\infty} e^{-t} t^{x-1} dt$$

be the Gamma function. Then:

- (i) $\Gamma(x)$ is convergent;
- (ii) $\Gamma(x+1) = x\Gamma(x)$ - in particular - $\Gamma(n+1) = n!$ for any $n \in \mathbb{N}$;
- (iii) $\lim_{x \rightarrow 0+} \Gamma(x) = \infty$.

For (i), we write

$$\int_{0+}^{\infty} e^{-t} t^{x-1} dt = \int_{0+}^1 e^{-t} t^{x-1} dt + \int_1^{\infty} e^{-t} t^{x-1} dt.$$

The first term is finite since $e^{-t} t^{x-1} \leq t^{x-1}$ for $t \in (0, 1)$ and $\int_{0+}^1 t^{x-1} dt = x^{-1}$ for $x > 0$. The second term is finite too due to the fact that $\lim_{t \rightarrow \infty} e^{-t} t^n = 0$ for all $n \in \mathbb{N}$, that $\int_1^{\infty} t^{-2} dt < \infty$, and that

$$e^{-t} t^{x-1} \leq e^{-t} t^{x+1} t^{-2} < t^{-2} \text{ for all } n \geq x+1.$$

For (ii) Let $\varepsilon, b > 0$. Using integration by parts, we have

$$\int_{\varepsilon}^1 e^{-t} t^{x-1} dt = -e^{-\varepsilon} \varepsilon^x x^{-1} + e^{-1} x^{-1} + x^{-1} \int_{\varepsilon}^1 e^{-t} t^x dt.$$

Letting $\varepsilon \rightarrow 0$, we get

$$\int_{0^+}^1 e^{-t} t^{x-1} dt = e^{-1} x^{-1} + x^{-1} \int_{0^+}^1 e^{-t} t^x dt.$$

Similarly,

$$\int_1^b e^{-t} t^{x-1} dt = e^{-b} b^x x^{-1} - e^{-1} x^{-1} + x^{-1} \int_1^b e^{-t} t^x dt$$

and letting $b \rightarrow \infty$, we obtain

$$\int_1^\infty e^{-t} t^{x-1} dt = -e^{-1} x^{-1} + x^{-1} \int_1^\infty e^{-t} t^x dt.$$

This implies $\Gamma(x) = x^{-1} \Gamma(x+1)$. Note also that $\Gamma(1) = 1 = 0!$. So the second formula follows from the first one via induction.

(iii) If $x > 0$ then

$$\Gamma(x) > \int_{0^+}^1 e^{-t} t^{x-1} dt \geq e^{-1} x^{-1} \rightarrow \infty \text{ as } x \rightarrow 0^+.$$

As an interesting application of some of the previous results, we introduce an important product – the so-called convolution of two functions, and use it to prove the polynomial approximation theorem on an interval.

Definition 2.4.9. Let $f_1, f_2 : \mathbf{R} \rightarrow \mathbf{R}$. The convolution $f_1 * f_2$ of f_1 and f_2 is defined by

$$f_1 * f_2(x) = \int_{-\infty}^{\infty} f_1(x-y) f_2(y) dy,$$

provided that the improper integral exists.

It is an easy exercise to prove that $f_1 * f_2 = f_2 * f_1$. Also, it can be shown that the convolution exists for a large class of functions; however, we shall restrict our discussion to the class of continuous functions.

Definition 2.4.10. A sequence $\phi_n : \mathbf{R} \rightarrow \mathbf{R}$ is a Dirac sequence or a delta sequence or an approximate identity provided that:

- (i) $\phi_n(t) \geq 0$ for all $t \in \mathbf{R}$;
- (ii) ϕ_n is continuous and integrable over \mathbf{R} with $\int_{-\infty}^{\infty} \phi_n(t) dt = 1$;

(iii) Given any $\delta > 0$, $\lim_{n \rightarrow \infty} \left(\int_{-\infty}^{-\delta} \phi_n(t) dt + \int_{\delta}^{\infty} \phi_n(t) dt \right) = 0$.

Example 2.4.11. $\phi_n(t) = (\pi(nt^2 + n^{-1}))^{-1}$ is the Siméon-Denis Poisson kernel. For our purpose, let us consider the Edmund Georg Hermann Landau kernel:

$$\phi_n(t) = \begin{cases} (1-t^2)^n c_n^{-1}, & |t| < 1 \\ 0, & |t| \geq 1 \end{cases} \quad \text{where } c_n = \int_{-1}^1 (1-t^2)^n dt.$$

It is obvious that this function satisfies (i) and (ii) of Definition 2.4.10. For (iii), note that

$$c_n = 2 \int_0^1 (1-t^2)^n dt \geq 2 \int_0^1 (1-t)^n dt = (n+1)^{-1}.$$

Thus, if $\delta \in (0, 1)$ is fixed, then

$$\int_{\delta}^1 \phi_n(t) dt = c_n^{-1} \int_{\delta}^1 (1-t^2)^n dt \leq 2^{-1}(n+1)(1-\delta^2)^n(1-\delta) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

By symmetry, $\int_{-1}^{-\delta} \phi_n(t) dt = \int_{\delta}^1 \phi_n(t) dt$, and (iii) follows.

Theorem 2.4.12. Let ϕ_n be a Dirac sequence, and $f : \mathbf{R} \rightarrow \mathbf{R}$ be bounded and continuous. Then $\lim_{n \rightarrow \infty} \sup_{x \in [a, b]} |f * \phi_n(x) - f(x)| = 0$ for any finite closed interval $[a, b]$ of \mathbf{R} .

Proof. The existence of $f * \phi_n$ is established by the following estimate that for any finite interval $[a, b]$,

$$\int_a^b |f(x-t)\phi_n(t)| dt \leq \sup_{x \in \mathbf{R}} |f(x)| \int_{-\infty}^{\infty} |\phi_n(t)| dt.$$

Using the continuity of f , we have that given a finite interval $[a, b]$ and for arbitrary $\varepsilon > 0$ there is a $\delta > 0$ such that $|f(x-t) - f(x)| < 2^{-1}\varepsilon$ for $x \in [a, b]$ and $|t| \leq \delta$. Since f is bounded, $M = \sup_{x \in \mathbf{R}} |f(x)| + 1$ is finite. Consequently, there is an $n_0 \in \mathbf{N}$ such that

$$n \geq n_0 \Rightarrow \int_{-\infty}^{-\delta} \phi_n(t) dt + \int_{\delta}^{\infty} \phi_n(t) dt < \varepsilon(4M)^{-1}.$$

Thus, for $x \in K$ and $n \geq n_0$ we get

$$\begin{aligned}
 |f * \phi_n(x) - f(x)| &= \left| \int_{-\infty}^{\infty} (f(x-t) - f(x)) \phi_n(t) dt \right| \\
 &\leq \left(\int_{-\infty}^{-\delta} + \int_{-\delta}^{\delta} + \int_{\delta}^{\infty} \right) |f(x-t) - f(x)| \phi_n(t) dt \\
 &\leq 2M \int_{-\infty}^{-\delta} \phi_n(t) dt + 2^{-1}\varepsilon \int_{-\delta}^{\delta} \phi_n(t) dt + 2M \int_{\delta}^{\infty} \phi_n(t) dt \\
 &\leq 2M\varepsilon(4M)^{-1} + 2^{-1}\varepsilon \int_{-\infty}^{\infty} \phi_n(t) dt \\
 &= \varepsilon.
 \end{aligned}$$

□

By choosing the Landau kernel, we obtain the classical Weierstrass approximation theorem for continuous functions on bounded and closed intervals on \mathbf{R} as follows.

Theorem 2.4.13. *Let $f \in C[a, b]$. Then for each $\varepsilon > 0$ there is a polynomial p such that $|f(x) - p(x)| < \varepsilon$ for all $x \in [a, b]$.*

Proof. We first simplify the problem by making some changes of variables. Let $u = (x - a)(b - a)^{-1}$. Then $x = (b - a)u + a$ and $u \in [0, 1]$. Set $g(u) = f((b - a)u + a) = f(x)$. If we can find a polynomial p such that $|p(u) - g(u)| < \varepsilon$ for all $u \in [0, 1]$, then

$$|p((x - a)(b - a)^{-1}) - f(x)| < \varepsilon$$

for all $x \in [a, b]$, and the function defined by $p((x - a)(b - a)^{-1})$ is a polynomial. So, we may assume that $[a, b] = [0, 1]$.

Next, let $h(x) = f(x) - f(0) - (f(1) - f(0))x$. If we can approximate h uniformly by a polynomial on $[0, 1]$, then f can be uniformly approximated by a polynomial on $[0, 1]$. Since $h(0) = h(1) = 0$, we may assume that $f(0) = f(1) = 0$.

Assuming this, we extend f continuously to \mathbf{R} by setting $f(x) = 0$ for $x \notin [0, 1]$. Then the extension, which we still denote by f , is bounded and continuous on \mathbf{R} . Suppose ϕ_n is the Landau kernel. It is not hard to see that $\phi_n(x - t) = (1 - (x - t)^2)^n c_n^{-1}$ is a polynomial in x and t ; so

$$\int_{-\infty}^{\infty} \phi_n(x - t) f(t) dt = \phi_n * f(x) = f * \phi_n(x)$$

is a polynomial in x . The result now follows from Theorem 2.4.12. □

Problems

2.1. Verify following four results.

(i) $f \in R[0, 1]$ and $\int_0^1 f(x) dx = 0$ for

$$f(x) = \begin{cases} 1, & x = 2^{-1} \\ 0, & x \neq 2^{-1} \end{cases}.$$

(ii) $f \in R[0, 1]$ and $\int_0^1 f(x) dx = 0$ for

$$f(x) = \begin{cases} 1, & x = 1, 2^{-1}, 3^{-1}, \dots \\ 0, & \text{otherwise} \end{cases}.$$

(iii) $f \in R[0, 2]$ and $\int_0^2 f(x) dx = -1$ for

$$f(x) = \begin{cases} 1, & x \in [0, 1] \\ -2, & x \in (1, 2] \end{cases}.$$

(iv) $f \in R[0, 2]$ and $\int_0^2 f(x) dx = 3$ for

$$f(x) = \begin{cases} 2, & x \in [0, 1) \\ 5, & x = 1 \\ 1, & x \in (1, 2] \end{cases}.$$

2.2. Prove that if $f \in R[0, 1]$ then

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n f(kn^{-1}) = \int_0^1 f(x) dx.$$

What about $f(x) = \sqrt{1-x^2}$?

2.3. Determine whether or not

$$f(x) = \begin{cases} 2, & x \leq 0 \\ x^{-1}, & x > 0 \end{cases}$$

is Riemann integrable on $[0, 1]$, $[-1, 0]$, and $[-1, 1]$ respectively.

2.4. Prove that if $f \in R[a, b]$ and $m \leq f(x) \leq M$ for all $x \in [a, b]$, then $m(b-a) \leq \int_a^b f(x) dx \leq M(b-a)$.

2.5. Given $f(x) = x$ and $P_n = \{kn^{-1}\}_{k=0}^n$, find $L(f, P_n)$ and $U(f, P_n)$. Then, show $f \in R[0, 1]$. Finally, evaluate $\int_0^1 f(x) dx$.

2.6. Prove that if $f \in R[a, b]$ then $|f| \in R[a, b]$ with $|\int_a^b f(x) dx| \leq \int_a^b |f(x)| dx$, but not conversely.

2.7. Let f be bounded on $[a, b]$. Prove that $f \in R[a, b]$ if and only if $f \in R[c, d]$ for all $[c, d] \subset (a, b)$ if and only if $f \in R[a, c]$ and $f \in R[c, b]$ for some $c \in (a, b)$.

2.8. Prove $\int_0^\pi \sqrt{x \sin x} dx \leq \pi$.

2.9. Let $f, g \in R[a, b]$. Prove the Minkowski inequality:

$$\left(\int_a^b (f(x) + g(x))^2 dx \right)^{2^{-1}} \leq \left(\int_a^b f^2(x) dx \right)^{2^{-1}} + \left(\int_a^b g^2(x) dx \right)^{2^{-1}}.$$

2.10. Show that if $f \in R[a, b]$ and there is a constant $\delta > 0$ such that $f(x) \geq \delta > 0$ for all $x \in [a, b]$, then $1/f \in R[a, b]$.

2.11. Prove that $f(x) = \sqrt{x}$ is uniformly continuous on $[0, \infty)$ by choosing $\delta = 2^{-2}\epsilon^2$.

2.12. Prove that if f is uniformly continuous on (a, b) then f is bounded there.

2.13. Prove that $f(x) = \cos(x^{-1})$ is not uniformly continuous on $(0, 1]$.

2.14. Prove that if $f(x)$ is the linear function $a_n x + b_n$ on $[n-1, n]$ for $n = 1, 2, \dots, k$ then $f \in R[0, k-2^{-1}]$.

2.15. Let

$$f(x) = \begin{cases} x^{-1} \sin x, & x \neq 0 \\ 3, & x = 0 \end{cases}.$$

Prove $f \in R[0, 5]$.

2.16. Prove that if $f \in R[a, b]$ and $F(x) = \int_a^x f(t) dt$ then $F \in C[a, b]$.

2.17. Prove that if $f \in C[a, b]$ with $f > 0$ and $F(x) = \int_a^x f(t) dt$ then F is strictly increasing on $[a, b]$; that is, $x_1 < x_2 \Rightarrow F(x_1) < F(x_2)$.

2.18. Let

$$f(x) = \begin{cases} \sin x^{-1}, & x \neq 0 \\ 0, & x = 0 \end{cases}.$$

Is f continuous on $[0, \pi^{-1}]$? Is f in $R[0, \pi^{-1}]$? Explain your answers.

2.19. For what values of p is $\int_1^\infty x^p dx$ or $\int_{0^+}^1 x^p dx$ convergent?

2.20. In the sequel, determine whether the improper integral is convergent:

(i) $\int_1^\infty (x\sqrt{x+1})^{-1} dx$;

(ii) $\int_{-\infty}^\infty xe^{-x^2} dx$;

(iii) $\int_0^{\frac{\pi}{2}-} \tan x dx$;

(iv) $\int_{1^+}^\infty (x(\ln x)^2)^{-1} dx$.

2.21. Prove that $\int_{-1}^1 x^{-1} dx$ and $\int_{-\infty}^\infty 2x(x^2+1)^{-1} dx$ are divergent, but their Cauchy principal values are finite:

$$\lim_{a \rightarrow 0^+} \left(\int_{-1}^{-a} x^{-1} dx + \int_a^1 x^{-1} dx \right) = 0 \quad \text{and} \quad \lim_{a \rightarrow \infty} \int_{-a}^a 2x(x^2+1)^{-1} dx = 0.$$

2.22. For $a, b > 0$ let $f(x) = e^{-a|x|}$ and $g(x) = e^{-b|x|}$. Evaluate $f * g$.

2.23. Let $f \in C[0, 1]$ and $\int_0^1 f(x)x^n dx = 0$ for $n = 0, 1, 2, \dots$. Prove $f = 0$.

2.24. Evaluate $\int_{0^+}^{1^-} (\ln u^{-1})^{-2^{-1}} du$.

2.25. Prove $\Gamma(n+2^{-1}) = (2n)! \sqrt{\pi} (4^n n!)^{-1}$ for any nonnegative integer n .

2.26. If the domain of f includes $[0, \infty)$, then the Pierre-Simon Laplace transform of f is defined by

$$L(f)(x) = \int_{0^+}^\infty e^{-xt} f(t) dt.$$

Note that if $f(t)$ is bounded as $t \rightarrow 0^+$, then $L(f)(x)$ is the improper integral of $e^{-xt} f(t)$ on $[0, \infty)$, and if $f(t)$ is unbounded as $t \rightarrow 0^+$, then we write

$$L(f)(x) = \int_{0^+}^b e^{-xt} f(t) dt + \int_b^\infty e^{-xt} f(t) dt.$$

Prove the following results for $a, b \in \mathbf{R}$:

(i) $L(e^{-x}) = (x+1)^{-1}$ for $x > -1$ and $L(x) = x^{-2}$ for $x > 0$;

(ii) $L(af + bg) = aL(f) + bL(g)$;

(iii) If $L(f)(x)$ is defined for $x > b$, then $L(e^{ax} f(x)) = L(f)(x-a)$ for $x > a+b$;

(iv) If f has a continuous derivative on $[0, \infty)$ and if $\lim_{t \rightarrow \infty} e^{-xt} f(t) = 0$ and $L(f)(x)$ exists when $x > a$, then $L(f')(x) = xL(f)(x) - f(0)$ for $x > a$;

(v) $L(xe^{ax}) = (x-a)^{-2}$ and $L(e^{ax} + axe^{ax}) = x(x-a)^{-2}$ for $x > a$.

Chapter 3

Riemann-Stieltjes Integrals

In this chapter, we work with the so-called Riemann-Stieltjes integration. This integration is Thomas Joannes Stieltjes' generalization of the Riemann integration. A particularly convenient feature of the Riemann-Stieltjes integration is that it can be employed to represent discrete sums.

3.1 Functions of Bounded Variation

Recall that a partition P of a closed interval $[a, b]$ of \mathbf{R} is such a set $\{x_k\}_{k=0}^n$ that $a = x_0 < x_1 < \cdots < x_n = b$. Let f be a function on $[a, b]$ and consider the sum

$$S(f, P) = \sum_{k=1}^n |f(x_k) - f(x_{k-1})|.$$

We wish to handle the set of all values of $S(f, P)$, where P ranges over all partitions of $[a, b]$. This set of values may be bounded above or unbounded above, and we write the supremum of this set of values as $\sup_P S(f, P)$.

Definition 3.1.1. *The function f is said to have bounded variation on $[a, b]$ provided that $\sup_P S(f, P)$ is finite. In this case $V_a^b f$ is used as $\sup_P S(f, P)$, the variation of f on $[a, b]$. Moreover, the collection of all functions having bounded variation on $[a, b]$ is written as $BV[a, b]$.*

Example 3.1.2. *If f is a monotonic function on $[a, b]$, then $S(f, P) = |f(b) - f(a)|$ for any partition P of $[a, b]$. Moreover, if*

$$f(x) = \begin{cases} 0, & x \leq 1 \\ 1, & x > 1 \end{cases} \quad \text{and} \quad g(x) = \begin{cases} 0, & x \neq 1 \\ 1, & x = 1 \end{cases},$$

then $V_0^2 f = 1$ and $V_0^2 g = 2$.

The first theorem on this topic is a simple, sufficient condition to ensure that a function has bounded variation.

Theorem 3.1.3. *If f has bounded derivative on $[a, b]$, then $f \in BV[a, b]$. Conversely, if $f \in BV[a, b]$, then f is bounded on $[a, b]$.*

Proof. First, let f' be bounded on $[a, b]$ and P be any partition $P = \{x_k\}_{k=0}^n$ of $[a, b]$. Then f is differentiable on the k th subinterval $[x_{k-1}, x_k]$, and hence by the mean value theorem for derivatives there exists a number $\xi_k \in [x_{k-1}, x_k]$ such that

$$f'(\xi_k)\Delta x_k = f(x_k) - f(x_{k-1}).$$

Using this, we get

$$S(f, P) = \sum_{k=1}^n |f'(\xi_k)|\Delta x_k \leq \sup_{x \in [a, b]} |f'(x)|(b-a).$$

Hence

$$V_a^b f \leq \sup_{x \in [a, b]} |f'(x)|(b-a).$$

Conversely, suppose $f \in BV[a, b]$. If $x \in [a, b]$, let P be the simple partition consisting of $\{a, x, b\}$. Then

$$S(f, P) = |f(x) - f(a)| + |f(b) - f(x)| \leq V_a^b f.$$

Therefore, $|f(x)| \leq |f(a)| + V_a^b f$, i.e., f is bounded on $[a, b]$. \square

Of course, the finiteness of $V_a^b f$ does not imply that f has bounded derivative. Next we see that differentiability does not imply bounded variation.

Example 3.1.4. *Let*

$$f(x) = \begin{cases} x^2 \sin \frac{\pi}{2x^2}, & x \neq 0 \\ 0 & , x = 0. \end{cases}$$

Then $f \notin BV[0, 1]$ although f is differentiable on $[0, 1]$.

To see $V_0^1 f = \infty$, for $n \in \mathbf{N}$ we choose a partition $P = \{x_j\}_{j=0}^n$ where

$$x_0 = 0, x_1 = \frac{1}{\sqrt{2n-1}}, \dots, x_k = \frac{1}{\sqrt{2(n-k)+1}}, \dots, x_{n-1} = \frac{1}{\sqrt{3}}, x_n = 1.$$

This choice gives

$$|f(x_{n-k}) - f(x_{n-k+1})| = \frac{1}{2k+1} + \frac{1}{2k-1} > \frac{2}{2k+1}.$$

Note that the series $\sum_{k=1}^{\infty} (2k+1)^{-1}$ is divergent. So it follows that by letting n be large enough, there will be enough terms in $S(f, P)$ so that the corresponding partial sum exceeds any given big number. To establish the differentiability of f , it suffices to consider this at $x = 0$. But

$$f'(0) = \lim_{x \rightarrow 0} x \sin \frac{\pi}{2x^2} = 0 \quad \text{since} \quad \left| x \sin \frac{\pi}{2x^2} \right| \leq |x|.$$

Remark 3.1.5. Suppose that $f \in BV[a, b]$ and $x \in [a, b]$. It is easy to show that $f \in BV[a, x]$ since any partition P of $[a, x]$ can be extended to a partition P^* of $[a, b]$ by adjoining the one additional point $x_{n+1} = b$. Clearly,

$$S(f, P) \leq S(f, P^*) \leq V_a^b f.$$

Thus we can define the variation function v_f by $v_f(x) = V_a^x f$ for $x \in [a, b]$.

Lemma 3.1.6. Let $f \in BV[a, b]$. Then:

- (i) $V_a^y f = V_a^x f + V_x^y f$ for any $a \leq x < y \leq b$;
- (ii) v_f and $v_f - f$ are increasing on $[a, b]$.

Proof. (i) Let $a \leq x < y \leq b$. We first show $V_a^y f \geq V_a^x f + V_x^y f$. In fact, if P_1 and P_2 are any partitions of $[a, x]$ and $[x, y]$ respectively, then $P_1 \cup P_2$ is a partition of $[a, y]$, and

$$\begin{aligned} S(f, P_1) + S(f, P_2) &= \sum_{x_k \leq x} |f(x_k) - f(x_{k-1})| + \sum_{x_k > x} |f(x_k) - f(x_{k-1})| \\ &= \sum_{x_k \in P_1 \cup P_2} |f(x_k) - f(x_{k-1})| \\ &= S(f, P_1 \cup P_2) \\ &\leq \sup\{S(f, P) : \text{all partitions } P \text{ of } [a, y]\}. \end{aligned}$$

Accordingly,

$$\sup_{P_1} S(f, P_1) + \sup_{P_2} S(f, P_2) \leq V_a^y f;$$

that is, $V_a^x f + V_x^y f \leq V_a^y f$.

Secondly, we prove the inverse inequality. To this end, let P be a partition of $[a, y]$ and P^* the partition obtained by adding the point x to P . Also, let P_1 and P_2 be the partitions of $[a, x]$ and $[x, y]$ respectively, induced by P^* . Then

$$S(f, P) \leq S(f, P^*) = S(f, P_1) + S(f, P_2) \leq V_a^x f + V_x^y f.$$

Taking the supremum over all partitions P yields $V_a^x f + V_x^y f \geq V_a^y f$. So, (i) follows.

(ii) That v_f is increasing on $[a, b]$ is clear from (i). Next, we prove that $v_f - f$ is increasing there. If $a \leq x \leq y \leq b$, then

$$(i) \Rightarrow V_a^y f - f(y) - (V_a^x f - f(x)) = V_x^y f - (f(y) - f(x)) \geq 0.$$

The reason why the right-hand number is nonnegative is that $|f(y) - f(x)|$ is precisely $S(f, P)$ when P is the trivial partition $\{x, y\}$. So, we conclude

$$v_f(x) - f(x) \leq v_f(y) - f(y),$$

as desired. □

Theorem 3.1.7. $f \in BV[a, b]$ if and only if there exist increasing functions f_1 and f_2 such that $f = f_1 - f_2$ on $[a, b]$.

Proof. Suppose that $f = f_1 - f_2$ on $[a, b]$, where f_1 and f_2 are increasing. Then $f_1, f_2 \in BV[a, b]$, so $f_1 - f_2 \in BV[a, b]$.

Conversely, if $f \in BV[a, b]$, then Lemma 3.1.6 shows that v_f and $v_f - f$ are increasing functions on $[a, b]$, and it is obvious that $f = v_f - (v_f - f)$. □

It is also the case that if f is a continuous function of bounded variation, then f can be written as the difference of two increasing continuous functions. From the proof of Theorem 3.1.7, this statement would follow immediately if the variation function v_f were continuous when f is continuous. This is precisely the content of the next theorem.

Theorem 3.1.8. Let $f \in BV[a, b]$. If f is continuous at $x_0 \in [a, b]$, then v_f is continuous at x_0 .

Proof. If $\varepsilon > 0$ and $x_0 < b$, then there is a partition P of $[x_0, b]$ such that

$$V_{x_0}^b f < S(f, P) + 2^{-1}\varepsilon.$$

Since f is continuous at x_0 , we may add a point x_1 to P to obtain a partition $P^* = \{x_0, x_1, \dots, x_n\}$ of $[x_0, b]$ such that $|f(x_1) - f(x_0)| < 2^{-1}\epsilon$. Then

$$S(f, P^*) = |f(x_1) - f(x_0)| + \sum_{j=2}^n |f(x_j) - f(x_{j-1})| < 2^{-1}\epsilon + V_{x_1}^b f$$

implies

$$V_{x_0}^b f < S(f, P) + 2^{-1}\epsilon \leq S(f, P^*) + 2^{-1}\epsilon < \epsilon + V_{x_1}^b f.$$

Accordingly, $0 \leq v_f(x_1) - v_f(x_0) < \epsilon$ holds due to Lemma 3.1.6 (ii). Since Lemma 3.1.6 (ii) indicates also that v_f is increasing, the limit $\lim_{x \rightarrow x_0^+} v_f(x)$ exists and by the last inequality, must equal $v_f(x_0)$.

Similarly, if $a < x_0$, then $\lim_{x \rightarrow x_0^-} v_f(x) = v_f(x_0)$, and hence v_f is continuous at x_0 . \square

3.2 Definition and Basic Properties

We start by defining the Riemann-Stieltjes integral on a finite closed interval.

Definition 3.2.1. Suppose that f and g are real-valued functions defined and bounded on $[a, b]$. Let $P = \{x_k\}_{k=0}^n$ be a partition of $[a, b]$ and $\{\xi_k\}_{k=1}^n$ be such that each ξ_k is in the k -th subinterval associated with P ; that is, $x_{k-1} \leq \xi_k \leq x_k$. Then

(i)

$$\sum_{k=1}^n f(\xi_k)(g(x_k) - g(x_{k-1}))$$

is called a Riemann-Stieltjes sum for f with respect to g .

(ii) f is called Riemann-Stieltjes integrable with respect to g on $[a, b]$ provided that there is a number J such that if $\epsilon > 0$ then there is a $\delta > 0$ such that

$$\|P\| < \delta \Rightarrow \left| J - \sum_{k=1}^n f(\xi_k)(g(x_k) - g(x_{k-1})) \right| < \epsilon$$

regardless of the choice of $\xi_k \in [x_{k-1}, x_k]$. In this case J is called the Riemann-Stieltjes integral of f with respect to g , and denoted by $\int_a^b f(x) dg(x)$. Also, f is called the integrand and g is called the integrator. The class of all functions Riemann-Stieltjes integrable on $[a, b]$ with respect to g is written as $RS_g[a, b]$.

Note that if $g(x) = x$ for $x \in [a, b]$ then the above definition goes back to the Riemann setting. However, the Riemann-Stieltjes integrals are different from the Riemann integrals.

Example 3.2.2. Suppose that $f \in C[a, b]$ and

$$g(x) = \begin{cases} 0, & x \in [a, t] \\ 1, & x \in (t, b] \end{cases}.$$

Given $P = \{x_k\}_{k=0}^n$, the point of discontinuity t is in either one or two subintervals: either $x_{m-1} < t < x_m$, or else $t = x_m$, which gives $[x_{m-1}, t]$ and $[t, x_{m+1}]$ as subintervals corresponding to P . All the terms of the Riemann-Stieltjes sum that do not involve these subintervals are 0, so in the first case the sum reduces to $f(\xi_m)$ and in the second case the sum reduces to $f(\xi_m) \cdot 0 + f(\xi_{m+1}) = f(\xi_{m+1})$. As $\|P\| \rightarrow 0$, both ξ_m and ξ_{m+1} approach t ; therefore the continuity of f implies that the limit of the Riemann-Stieltjes sums exists and equals $f(t)$. Hence $\int_a^b f(x) dg(x) = f(t)$.

The forthcoming theorem gives the connection between the Riemann-Stieltjes integrals and the Riemann integrals suggested by the use of the differential notation dg :

Theorem 3.2.3. Let $f \in R[a, b]$ and $g' \in C[a, b]$. Then

$$f \in RS_g[a, b] \text{ and } \int_a^b f(x) dg(x) = \int_a^b f(x) g'(x) dx.$$

Proof. Given any partition $P = \{x_k\}_{k=0}^n$ of $[a, b]$, g is differentiable on the k th subinterval $[x_{k-1}, x_k]$, so by the mean value theorem for derivatives there exists a number $c_k \in [x_{k-1}, x_k]$ such that

$$g'(c_k) \Delta x_k = g(x_k) - g(x_{k-1}).$$

Using this, we get

$$\sum_{k=1}^n f(\xi_k) (g(x_k) - g(x_{k-1})) = \sum_{k=1}^n f(\xi_k) g'(\xi_k) \Delta x_k + \sum_{k=1}^n f(\xi_k) (g'(c_k) - g'(\xi_k)) \Delta x_k.$$

The third sum in the last equation is a Riemann sum for $f g' \in R[a, b]$. Then, as $\|P\| \rightarrow 0$ this sum approaches $\int_a^b f(x) g'(x) dx$. Hence we can complete our proof by showing that the second sum of the last equation approaches 0 as $\|P\| \rightarrow 0$.

To achieve this goal we may assume $M = \sup_{x \in [a, b]} |f(x)| \neq 0$ and refer to the uniform continuity of g' – for $\varepsilon > 0$ choose $\delta > 0$ so that

$$|\xi - c| < \delta \Rightarrow |g'(c) - g'(\xi)| < \varepsilon(b-a)^{-1}M^{-1}.$$

Now,

$$\|P\| < \delta \Rightarrow |c_k - \xi_k| < \delta,$$

and consequently,

$$\begin{aligned} \left| \sum_{k=1}^n f(\xi_k)(g'(c_k) - g'(\xi_k))\Delta x_k \right| &\leq \sum_{k=1}^n |f(\xi_k)| |g'(c_k) - g'(\xi_k)| \Delta x_k \\ &< (b-a)^{-1}M^{-1}\varepsilon \max_{1 \leq k \leq n} |f(\xi_k)| \sum_{k=1}^n \Delta x_k \\ &\leq \varepsilon. \end{aligned}$$

Thus, as $\|P\| \rightarrow 0$, the Riemann-Stieltjes sums approach $\int_a^b f(x)g'(x)dx$. \square

The algebraic closure of the Riemann-Stieltjes integrals is included in the formulas (i) below.

Theorem 3.2.4. *Let $f_1, f_2 \in RS_{g_1}[a, b] \cap RS_{g_2}[a, b]$.*

(i) *If $c_1, c_2 \in \mathbf{R}$, then*

$$\begin{cases} \int_a^b (c_1 f_1(x) + c_2 f_2(x)) dg_1(x) = c_1 \int_a^b f_1(x) dg_1(x) + c_2 \int_a^b f_2(x) dg_1(x); \\ \int_a^b f_1(x) d(c_1 g_1(x) + c_2 g_2(x)) = c_1 \int_a^b f_1(x) dg_1(x) + c_2 \int_a^b f_1(x) dg_2(x). \end{cases}$$

(iii) *If $f_1 \leq f_2$ and g is increasing on $[a, b]$, then*

$$\int_a^b f_1(x) dg(x) \leq \int_a^b f_2(x) dg(x).$$

Proof. This is left as an exercise. \square

The foregoing result is based on the condition: $a \leq b$. In order to consider an integral such as $\int_b^a f(x) dg(x)$, where $a < b$, we adopt the familiar convention

$$\int_b^a f(x) dg(x) = - \int_a^b f(x) dg(x).$$

Theorem 3.2.5. *Let $f \in RS_g[a, b] \cap RS_g[b, c] \cap RS_g[a, c]$. Then*

$$\int_a^c f(x) dg(x) = \int_a^b f(x) dg(x) + \int_b^c f(x) dg(x).$$

Proof. It suffices to prove the formula holds under the assumption $a < b < c$. Given $\epsilon > 0$, choose $\delta > 0$ so that if P is a partition of $[a, b]$ or $[b, c]$ or $[a, c]$ satisfying $\|P\| < \delta$, then any choice of ξ_k 's in those subintervals induced by P yields a Riemann-Stieltjes sum, written as $\sum_a^b(\dots)$ or $\sum_b^c(\dots)$ or $\sum_a^c(\dots)$ which is within $3^{-1}\epsilon$ of the Riemann-Stieltjes integral, denoted by $\int_a^b(\dots)$ or $\int_b^c(\dots)$ or $\int_a^c(\dots)$. So,

$$\|P\| < \delta \Rightarrow \left| \int_a^b(\dots) - \sum_a^b(\dots) \right| + \left| \int_b^c(\dots) - \sum_b^c(\dots) \right| + \left| \int_a^c(\dots) - \sum_a^c(\dots) \right| < 3^{-1}\epsilon.$$

Since $a < b < c$, the sum $\sum_a^b(\dots) + \sum_b^c(\dots)$ becomes the Riemann-Stieltjes sum $\sum_a^c(\dots)$ for $[a, c]$. Thus we can write

$$\begin{aligned} & \left| \int_a^c(\dots) - \left(\int_a^b(\dots) + \int_b^c(\dots) \right) \right| \\ &= \left| \left(\int_a^c(\dots) - \sum_a^c(\dots) \right) - \left(\int_a^b(\dots) - \sum_a^b(\dots) \right) - \left(\int_b^c(\dots) - \sum_b^c(\dots) \right) \right| \\ &\leq \left| \int_a^c(\dots) - \sum_a^c(\dots) \right| + \left| \int_a^b(\dots) - \sum_a^b(\dots) \right| + \left| \int_b^c(\dots) - \sum_b^c(\dots) \right| \\ &< 3^{-1}\epsilon + 3^{-1}\epsilon + 3^{-1}\epsilon \\ &= \epsilon. \end{aligned}$$

Namely,

$$\left| \int_a^c f(x) dg(x) - \left(\int_a^b f(x) dg(x) + \int_b^c f(x) dg(x) \right) \right| < \epsilon.$$

Consequently, the desired formula follows. \square

It is worth pointing out that in the above theorem we assume that all three integrals exist and then conclude that the formula holds. In the Riemann integral analogue, it was sufficient to assume only the existence of $\int_a^b(\dots)$ and $\int_b^c(\dots)$ which implies the existence of $\int_a^c(\dots)$. That implication is not valid for the Riemann-Stieltjes integral, as we see in the following example:

Example 3.2.6. *If*

$$f(x) = \begin{cases} 0, & x \in [0, 1] \\ 1, & x \in (1, 2] \end{cases} \quad \text{and} \quad g(x) = \begin{cases} 0, & x \in [0, 1) \\ 1, & x \in [1, 2] \end{cases},$$

then $\int_0^1 f(x) dg(x) = 0$ and $\int_1^2 f(x) dg(x) = 0$. But $\int_0^2 f(x) dg(x)$ does not exist since f and g are discontinuous at 1; see the forthcoming Theorem 3.3.2.

3.3 Nonexistence and Existence for Integrals

We now proceed to establish some results on the nonexistence and existence of Riemann-Stieltjes integrals for special functions such as discontinuous functions, step functions and continuous functions.

Example 3.3.1. *If*

$$f(x) = g(x) = \begin{cases} 0, & x \in [0, 1] \\ 1, & x \in (1, 2] \end{cases},$$

then $f \notin RS_g[0, 2]$. In fact, if P is a partition of $[0, 2]$ having an arbitrarily small norm and $x_{m-1} = 1 \in P$, then $g(x_m) - g(x_{m-1}) = 1$, and hence

$$\sum_{k=1}^n f(\xi_k)(g(x_k) - g(x_{k-1})) = \begin{cases} 0, & \xi_m = 1 \\ 1, & \xi_m > 1 \end{cases}.$$

Examples 3.2.6 and 3.3.1 illustrate a general non-existence result as follows.

Theorem 3.3.2. *If $f, g : [a, b] \rightarrow \mathbf{R}$ are both discontinuous at $c \in [a, b]$, then $f \notin RS_g[a, b]$.*

Proof. Two cases will be taken into account. First suppose that $\lim_{x \rightarrow c} g(x)$ does not exist. Then there is an $\epsilon_g > 0$ such that for any $\delta > 0$ we can choose $x_{m-1}, x_m \in [a, b]$ satisfying

$$x_{m-1} < c < x_m; \quad x_m - x_{m-1} < \delta; \quad |g(x_m) - g(x_{m-1})| \geq \epsilon_g.$$

Now let P be a partition of $[a, b]$ with $\|P\| < \delta$ whose m th subinterval is $[x_{m-1}, x_m]$. The discontinuity of f at c implies the existence of an $\epsilon_f > 0$ and $\xi_{m,1}, \xi_{m,2} \in [x_{m-1}, x_m]$ such that

$$|f(\xi_{m,1}) - f(\xi_{m,2})| \geq \epsilon_f.$$

If we choose $\xi_{k,1} = \xi_{k,2}$ for $k \neq m$, then the corresponding Riemann-Stieltjes sums differ by at least $\varepsilon_f \varepsilon_g$. Since $\|P\|$ is arbitrarily small, this shows that the Riemann-Stieltjes sums cannot approach a finite number as $\|P\| \rightarrow 0$.

Next, we consider the case in which $\lim_{x \rightarrow c} g(x)$ exists but does not equal $g(c)$, say,

$$|g(c) - \lim_{x \rightarrow c} g(x)| = \varepsilon_g > 0.$$

For a given $\delta > 0$, choose a partition P of $[a, b]$ such that $\|P\| < \delta$, $x_m = c$, and

$$\text{either } |g(c) - g(x_{m+1})| \geq 2^{-1} \varepsilon_g \text{ or } |g(c) - g(x_{m-1})| \geq 2^{-1} \varepsilon_g.$$

The discontinuity of f at c implies the existence of an $\varepsilon_f > 0$ such that either $[x_{m-1}, x_m]$ or $[x_m, x_{m+1}]$ contains points $\xi_{m,1}$ and $\xi_{m,2}$ such that

$$|f(\xi_{m,1}) - f(\xi_{m,2})| \geq \varepsilon_f.$$

As above, this produces two Riemann-Stieltjes sums for P that differ by at least $2^{-1} \varepsilon_f \varepsilon_g$, so the sums do not approach a finite number as $\|P\| \rightarrow 0$. \square

Motivated by Examples 3.2.6 and 3.3.1, we consider integrators as step functions. These are functions which are constant on an interval except for a finite number of jump discontinuities.

Definition 3.3.3. Let g be defined on $[a, b]$ in such a way that g is discontinuous at a finite number of points $\{c_k\}_{k=1}^l$, where $l \in \mathbf{N}$ and $a \leq c_1 < c_2 < \cdots < c_l \leq b$. If g is a constant λ_k on each subinterval I_k with endpoints c_{k-1} and c_k , then g is called a step function and the number $g(c_k^+) - g(c_k^-)$ is called the jump of g at c_k , and hence $g = \sum_{k=1}^l \lambda_k 1_{I_k}$ where

$$1_{I_k}(x) = \begin{cases} 1, & x \in I_k \\ 0, & x \notin I_k \end{cases}.$$

If $c_1 = a$ and $c_l = b$, then the jumps of g at c_1 and c_l are $g(c_1^+) - g(c_1)$ and $g(c_l) - g(c_l^-)$ respectively. Here for each $k = 1, 2, \dots, n$,

$$g(c_k^+) = \lim_{0 < x - c_k \rightarrow 0} g(x) \text{ and } g(c_k^-) = \lim_{0 > x - c_k \rightarrow 0} g(x).$$

Step functions not only have bounded variation but also can be used to represent Riemann-Stieltjes integrals as finite sums.

Theorem 3.3.4. Let g be a step function on $[a, b]$ with jumps d_k at c_k where $p \in \mathbf{N}$ and $a \leq c_1 < c_2 < \cdots < c_l \leq b$. If $f : [a, b] \rightarrow \mathbf{R}$ is continuous at each c_k . Then $f \in RS_g[a, b]$ and

$$\int_a^b f(x) dg(x) = \sum_{k=1}^l f(c_k) d_k.$$

Proof. Let $\delta = \min_{1 \leq k \leq l} (c_k - c_{k-1})$. It is clear that for any partition $P = \{x_k\}_{k=0}^n$ satisfying $\|P\| < \delta$, each interval $I_k = [x_{k-1}, x_k]$ can contain at most one of the points $\{c_k\}_{k=1}^l$. If I_k contains no c_j then $g(x_k) - g(x_{k-1}) = 0$. If I_k contains the point c_j in its interior then $g(x_k) - g(x_{k-1}) = d_j$. Finally if the point c_j is common to $[x_{k-1}, x_k]$ and $[x_k, x_{k+1}]$; that is, if $c_j = x_k$, then

$$g(x_k) - g(x_{k-1}) + g(x_{k+1}) - g(x_k) = d_j.$$

Let $I_{k_1}, I_{k_2}, \dots, I_{k_r}$ ($k_1 < \cdots < k_r$) denote those intervals which contain a c_j . Then

$$\sum_{k=1}^n f(\xi_k) (g(x_k) - g(x_{k-1})) = \sum_{m=1}^r f(\xi_{k_m}) (g(x_{k_m}) - g(x_{k_{m-1}})).$$

If c_j is in the interior of I_{k_m} , then the continuity of f at c_j implies that

$$f(\xi_{k_m}) (g(x_{k_m}) - g(x_{k_{m-1}})) \rightarrow f(c_j) d_j \text{ as } \|P\| \rightarrow 0.$$

If c_j is common to I_{k_m} and $I_{k_{m+1}}$, then the continuity of f at c_j implies that

$$f(\xi_{k_m}) (g(x_{k_m}) - g(x_{k_{m-1}})) + f(\xi_{k_{m+1}}) (g(x_{k_{m+1}}) - g(x_{k_m})) \rightarrow f(c_j) d_j \text{ as } \|P\| \rightarrow 0.$$

Hence

$$\begin{aligned} \lim_{\|P\| \rightarrow 0} \sum_{k=1}^n f(\xi_k) (g(x_k) - g(x_{k-1})) &= \lim_{\|P\| \rightarrow 0} \sum_{m=1}^r f(\xi_{k_m}) (g(x_{k_m}) - g(x_{k_{m-1}})) \\ &= \sum_{k=1}^l f(c_k) d_k, \end{aligned}$$

establishing the desired formula. □

Remark 3.3.5. If in Theorem 3.3.4 one has $[a, b] = [1, n]$ for any $n - 1 \in \mathbf{N}$ and $g(x) = [x]$ - the greatest integer which is less than or equal to x , then for any real

numbers a_1, a_2, \dots, a_n one can choose a function $f \in C[1, n]$ such that $f(k) = a_k$ for each k and hence by Theorem 3.3.4,

$$\sum_{k=1}^n a_k = \sum_{k=1}^n f(k) = \int_0^n f(x) d[x].$$

Namely, a finite sum can be written as a Riemann-Stieltjes integral.

In general, we have the following existence theorem for Riemann-Stieltjes integrals regarding continuous functions.

Theorem 3.3.6. *Let $f \in C[a, b]$ and $g \in BV[a, b]$. Then $f \in RS_g[a, b]$.*

Proof. By Theorem 3.1.7, we know that $g = g_1 - g_2$, where g_1, g_2 are increasing. If we prove that $\int_a^b f(x) dg_1(x)$ exists for the increasing g_1 , then $\int_a^b f(x) dg_2(x)$ exists also for the increasing g_2 . This implies the existence of

$$\int_a^b f(x) d(g_1(x) - g_2(x)) = \int_a^b f(x) dg(x).$$

For the sake of simplifying notation, we write g in place of g_1 .

For any partition $P = \{x_k\}_{k=0}^n$, let

$$\begin{cases} I_k = [x_{k-1}, x_k]; \\ m_k = \inf\{f(x) : x \in I_k\}; \\ M_k = \sup\{f(x) : x \in I_k\}. \end{cases}$$

We also define the upper and lower sums of f with respect to g by

$$U(f, P; g) = \sum_{k=1}^n M_k (g(x_k) - g(x_{k-1})) \quad \text{and} \quad L(f, P; g) = \sum_{k=1}^n m_k (g(x_k) - g(x_{k-1})).$$

If $\xi_k \in I_k$, then $m_k \leq f(\xi_k) \leq M_k$, and hence

$$L(f, P; g) \leq \sum_{k=1}^n f(\xi_k) (g(x_k) - g(x_{k-1})) \leq U(f, P; g).$$

If P and P^* are two partitions of $[a, b]$, then $P \cup P^*$ is a refinement of both. It is easy to verify that refining a partition increases its lower sums and decreases its upper sums, so it follows that

$$L(f, P; g) \leq L(f, P \cup P^*; g) \leq U(f, P \cup P^*; g) \leq U(f, P^*; g);$$

therefore any lower sum is less than or equal to any upper sum. It follows that the least upper bound of the set of all lower sums cannot exceed the greatest lower bound of the set of all upper sums:

$$L(f; g) = \sup_P L(f, P; g) \leq \inf_P U(f, P; g) = U(f; g).$$

We can prove that $\int_a^b f(x) dg(x)$ exists if and only if for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\|P\| < \delta \Rightarrow U(f, P; g) - L(f, P; g) < \varepsilon,$$

or equivalently, for any $\varepsilon > 0$ there exists a partition P of $[a, b]$ such that $U(f, P; g) - L(f, P; g) < \varepsilon$. In fact, the proof is virtually identical to the proof of Theorem 2.2.5, but with $g(x_k) - g(x_{k-1})$ replacing $\Delta x_k = x_k - x_{k-1}$.

Because any Riemann-Stieltjes sum (using P) is between $L(f, P; g)$ and $U(f, P; g)$, and $L(f; g)$ is also between $L(f, P; g)$ and $U(f, P; g)$, it follows that any such Riemann-Stieltjes sum must also be within ε of $L(f; g)$. Thus we establish that the Riemann-Stieltjes sums converge to $L(f; g)$ as $\|P\| \rightarrow 0$. By the uniform continuity of $f \in C[a, b]$, we choose $\delta > 0$ so that for any $x, y \in [a, b]$,

$$|x - y| < \delta \Rightarrow |f(x) - f(y)| < 2^{-1}\varepsilon(g(b) - g(a))^{-1}.$$

Here, we assume naturally $g(a) \neq g(b)$, for otherwise the monotonic function g would be a constant and the conclusion is trivial. If $\|P\| < \delta$, then for $k = 1, \dots, n$,

$$M_k - m_k < \varepsilon(g(b) - g(a))^{-1}.$$

Therefore

$$\begin{aligned} U(f, P; g) - L(f, P; g) &= \sum_{k=1}^n (M_k - m_k)(g(x_k) - g(x_{k-1})) \\ &< \varepsilon(g(b) - g(a))^{-1} \sum_{k=1}^n (g(x_k) - g(x_{k-1})) \\ &< \varepsilon. \end{aligned}$$

□

The foregoing theorem can be used to handle composition with on Riemann-Stieltjes integrable functions.

Theorem 3.3.7. *Let g be increasing on $[a, b]$ and $f \in RS_g[a, b]$ with the range of f contained in the closed interval $[c, d]$. If $\psi \in C[c, d]$ then $\psi \circ f \in RS_g[a, b]$.*

Proof. Since $\psi \in C[c, d]$, it follows that

$$M = 1 + \sup_{x \in [c, d]} |\psi(x)| < \infty,$$

ψ is uniformly continuous on $[c, d]$, and consequently, for any $\varepsilon > 0$ there is a

$$\delta \in \left(0, \frac{\varepsilon}{g(b) - g(a) + 2M}\right)$$

such that

$$y_1, y_2 \in [c, d] \text{ and } |y_1 - y_2| < \delta \Rightarrow |\psi(y_1) - \psi(y_2)| < \frac{\varepsilon}{g(b) - g(a) + 2M}.$$

Because of $f \in RS_g[a, b]$ we can use the argument in Theorem 3.3.6 to obtain a partition $P = \{x_k\}_{k=0}^n$ of $[a, b]$ such that

$$U(f, P; g) - L(f, P; g) = \sum_{k=1}^n (M_k - m_k) (g(x_k) - g(x_{k-1})) \leq \delta^2.$$

If

$$M_k^* = \sup_{x \in [x_{k-1}, x_k]} \psi \circ f(x) \text{ and } m_k^* = \inf_{x \in [x_{k-1}, x_k]} \psi \circ f(x),$$

then

$$\begin{aligned} & \sum_{k=1}^n (M_k^* - m_k^*) (g(x_k) - g(x_{k-1})) \\ &= \left(\sum_{M_k - m_k < \delta} + \sum_{M_k - m_k \geq \delta} \right) (\cdots) \\ &\leq \frac{\varepsilon}{g(b) - g(a) + 2M} \sum_{k=1}^n (g(x_k) - g(x_{k-1})) + 2M \sum_{M_k - m_k \geq \delta} (g(x_k) - g(x_{k-1})) \\ &\leq \frac{\varepsilon(g(b) - g(a))}{g(b) - g(a) + 2M} + 2M\delta^{-1} \sum_{M_k - m_k \geq \delta} \delta (g(x_k) - g(x_{k-1})) \\ &\leq \frac{\varepsilon(g(b) - g(a))}{g(b) - g(a) + 2M} + 2M\delta^{-1} \sum_{k=1}^n (M_k - m_k) (g(x_k) - g(x_{k-1})) \\ &\leq \frac{\varepsilon(g(b) - g(a))}{g(b) - g(a) + 2M} + 2M\delta \\ &< \varepsilon, \end{aligned}$$

and hence $\psi \circ f \in RS_g[a, b]$. \square

Corollary 3.3.8. *Let g be increasing on $[a, b]$. If $f_1, f_2 \in RS_g[a, b]$, then*

$$f_1 f_2, |f_1|, |f_2|, \max\{f_1, f_2\}, \min\{f_1, f_2\} \in RS_g[a, b].$$

Proof. Letting $\psi(y) = y^2$ respectively $\psi(y) = |y|$ in Theorem 3.3.7, as well as noticing three identities:

$$\begin{cases} f_1 f_2 = 2^{-2}((f_1 + f_2)^2 - (f_1 - f_2)^2); \\ \max\{f_1, f_2\} = 2^{-1}(f_1 + f_2 + |f_1 - f_2|); \\ \min\{f_1, f_2\} = 2^{-1}(f_1 + f_2 - |f_1 - f_2|), \end{cases}$$

we obtain the desired result. \square

In what follows, we will see that in many cases the continuity is essential for the existence of a Riemann-Stieltjes integral.

Definition 3.3.9. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. A subset E of \mathbf{R} is called a g -null set if, for any $\varepsilon > 0$ there exists a sequence $\{(a_j, b_j)\}_{j=1}^{\infty}$ of open intervals such that*

$$E \subseteq \bigcup_{j=1}^{\infty} (a_j, b_j) \text{ and } \sum_{j=1}^{\infty} (g(b_j^-) - g(a_j^+)) < \varepsilon.$$

Remark 3.3.10. *Needless to say, if $g : \mathbf{R} \rightarrow \mathbf{R}$ is increasing, then the empty set is g -null; a subset of a g -null set is itself a g -null set; and the singleton $\{x\} \subseteq \mathbf{R}$ is a g -null set when and only when g is continuous at x .*

Theorem 3.3.11. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing continuous and $f : [a, b] \rightarrow \mathbf{R}$ bounded function. Then $f \in RS_g[a, b]$ when and only when the set of discontinuous points of f is a g -null set.*

Proof. Since f is bounded on $[a, b]$, we may always assume that

$$M = 1 + \sup_{x \in [a, b]} |f(x)| < \infty$$

but also that $\mathcal{D}(f)$ is the set of discontinuity points of f in $[a, b]$. For any interval $J \subseteq \mathbf{R}$ let

$$\omega_f(J) = \sup\{f(x) : x \in J \cap [a, b]\} - \inf\{f(x) : x \in J \cap [a, b]\},$$

and for $x \in [a, b]$ set

$$\omega_f(x) = \inf\{\omega_f(J) : J \text{ any open interval containing } x\}.$$

It is clear that f is continuous at x if and only if $\omega_f(x) = 0$. This result indicates

$$\mathcal{D}(f) = \cup_{k=1}^{\infty} \mathcal{D}_k(f) \text{ where } \mathcal{D}_k(f) = \{x \in [a, b] : \omega_f(x) \geq k^{-1}\}.$$

Because g is continuous on \mathbf{R} , one has $g(c^-) = g(c) = g(c^+)$ for $c \in \mathbf{R}$ – this fact plays an important role in the sequel.

On the one hand, suppose that f belongs to $RS_g[a, b]$. We want to prove that $\mathcal{D}(f)$ is a g -null set. Observe that if each $\mathcal{D}_k(f)$ is a g -null set then any $\varepsilon > 0$ there is a sequence of open intervals $\{(a_{k,j}, b_{k,j})\}_{j=1}^{\infty}$ such that

$$\mathcal{D}_k(f) \subseteq \cup_{j=1}^{\infty} (a_{k,j}, b_{k,j}) \text{ and } \sum_{j=1}^{\infty} (g(b_{k,j}) - g(a_{k,j})) < \varepsilon 2^{-k};$$

but then

$$\mathcal{D}(f) \subseteq \cup_{k=1}^{\infty} \cup_{j=1}^{\infty} (a_{k,j}, b_{k,j}) \text{ and } \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} (g(b_{k,j}) - g(a_{k,j})) < \varepsilon,$$

and so $\mathcal{D}(f)$ is a g -null set. Therefore, it suffices to verify that each $\mathcal{D}_k(f)$ is a g -null set. Fix now $k \in \mathbf{N}$. Because of $f \in RS_g[a, b]$, for any $\varepsilon > 0$ there are a partition $P = \{x_j\}_{j=0}^n$ of $[a, b]$ and the lower and upper Riemann-Stieltjes sums $L(f, P; g)$ and $U(f, P; g)$ (see also the argument for Theorem 3.3.6) such that $U(f, P; g) - L(f, P; g) < k^{-1}\varepsilon$. Consequently,

$$\begin{aligned} (2k)^{-1}\varepsilon &> \sum_{(x_{j-1}, x_j) \cap \mathcal{D}_k(f) \neq \emptyset} (M_j - m_j)(g(x_j) - g(x_{j-1})) \\ &+ \sum_{(x_{j-1}, x_j) \cap \mathcal{D}_k(f) = \emptyset} (M_j - m_j)(g(x_j) - g(x_{j-1})). \end{aligned}$$

Note that every point of $\mathcal{D}_k(f)$ belongs to P or satisfies

$$(x_{j-1}, x_j) \cap \mathcal{D}_k(f) \neq \emptyset \text{ for some } j \in \{1, \dots, n\}.$$

Accordingly,

$$\begin{aligned} (2k)^{-1}\varepsilon &> \sum_{(x_{j-1}, x_j) \cap \mathcal{D}_k(f) \neq \emptyset} (M_j - m_j)(g(x_j) - g(x_{j-1})) \\ &\geq k^{-1} \sum_{(x_{j-1}, x_j) \cap \mathcal{D}_k(f) \neq \emptyset} (g(x_j) - g(x_{j-1})), \end{aligned}$$

which means

$$\sum_{(x_{j-1}, x_j) \cap \mathcal{D}_k(f) \neq \emptyset} (g(x_j) - g(x_{j-1})) < 2^{-1} \varepsilon.$$

Since g is uniformly continuous on $[a-1, b+1]$, there exists a $\delta > 0$ such that

$$\sup_{0 \leq j \leq n} (g(x_j + \delta) - g(x_j - \delta)) < 2^{-1} \varepsilon (1+n)^{-1}.$$

But then

$$\mathcal{D}_k(f) \subseteq \left(\bigcup_{(x_{j-1}, x_j) \cap \mathcal{D}_k(f) \neq \emptyset} (x_{j-1}, x_j) \right) \left(\bigcup_{j=0}^n (x_j - \delta, x_j + \delta) \right)$$

and

$$\sum_{(x_{j-1}, x_j) \cap \mathcal{D}_k(f) \neq \emptyset} (g(x_j) - g(x_{j-1})) + \sum_{j=0}^n (g(x_j + \delta) - g(x_j - \delta)) < \varepsilon.$$

In other words, $\mathcal{D}_k(f)$ is a g -null set.

On the other hand, suppose that $\mathcal{D}(f)$ is a g -null set. We are about to prove $f \in RS_g[a, b]$. Thanks to $\mathcal{D}_k(f) \subseteq \mathcal{D}(f)$ for each $k \in \mathbf{N}$, we can readily derive that each $\mathcal{D}_k(f)$ is a g -null set too. Consequently, we can find open intervals $\{(a_{k,j}, b_{k,j})\}_{j=1}^\infty$ such that

$$\mathcal{D}_k(f) \subseteq \bigcup_{j=1}^\infty (a_{k,j}, b_{k,j}) \quad \text{and} \quad \sum_{j=1}^\infty (g(b_{k,j}) - g(a_{k,j})) < \varepsilon (4M)^{-1}.$$

Note that each $\mathcal{D}_k(f)$ is closed and bounded. So by the Heine-Borel property, we get a finite subset $\{(a_{k,j_l}, b_{k,j_l})\}_{l=1}^{n_0}$ of $\{(a_{k,j}, b_{k,j})\}_{j=1}^\infty$ such that $\mathcal{D}_k(f) \subseteq \bigcup_{l=1}^{n_0} (a_{k,j_l}, b_{k,j_l})$. Thus, $[a, b] \setminus \bigcup_{l=1}^{n_0} (a_{k,j_l}, b_{k,j_l})$ is a disjoint union of finitely many intervals $\{[c_{k,l}, d_{k,l}]\}_{l=1}^{l_0}$. Hence

$$[a, b] = \left(\bigcup_{l=1}^{l_0} [c_{k,l}, d_{k,l}] \right) \cup \left(\bigcup_{l=1}^{n_0} (a_{k,j_l}, b_{k,j_l}) \right).$$

Since $\omega_f(x) \geq k^{-1}$ whenever $x \in \mathcal{D}_k(f)$, we find that $\omega_f(x) < k^{-1}$ whenever $x \in \bigcup_{l=1}^{l_0} [c_{k,l}, d_{k,l}]$.

Now the step functions G and H are defined by

$$G(x) = \begin{cases} \inf_{t \in [c_{k,l}, d_{k,l}]} f(t), & x \in [c_{k,l}, d_{k,l}], \quad l = 1, \dots, l_0 \\ \inf_{t \in (a_{k,j_l}, b_{k,j_l})} f(t), & x \in (a_{k,j_l}, b_{k,j_l}), \quad l = 1, \dots, n_0 \end{cases},$$

and

$$H(x) = \begin{cases} \sup_{t \in [c_{k,l}, d_{k,l}]} f(t), & x \in [c_{k,l}, d_{k,l}], \quad l = 1, \dots, l_0 \\ \sup_{t \in (a_{k,j_l}, b_{k,j_l})} f(t), & x \in (a_{k,j_l}, b_{k,j_l}), \quad l = 1, \dots, n_0 \end{cases}.$$

Then, for any $\varepsilon > 0$ there is an $\check{n}_0 \in \mathbf{N}$ such that $k > \check{n}_0$ implies

$$\begin{aligned} & \int_a^b (H(x) - G(x)) dg(x) \\ & \leq k^{-1} \sum_{l=1}^{l_0} (g(d_{k,l}) - g(c_{k,l})) + 2M \sum_{l=1}^{\check{n}_0} (g(b_{k,j_l}) - g(a_{k,j_l})) \\ & \leq k^{-1} (g(b) - g(a)) + 2^{-1} \varepsilon \\ & < \varepsilon. \end{aligned}$$

Hence $f \in RS_g[a, b]$, due to the argument for Theorem 3.3.6. \square

3.4 Evaluations of Integrals

With the help of the previous discussion on the nonexistence and existence of Riemann-Stieltjes integrals, we can introduce some methods and techniques to evaluate Riemann-Stieltjes integrals.

Firstly, we compute each Riemann-Stieltjes integral by parts.

Theorem 3.4.1. *If $f \in RS_g[a, b]$, then $g \in RS_f[a, b]$ and*

$$\int_a^b f(x) dg(x) + \int_a^b g(x) df(x) = f(b)g(b) - f(a)g(a).$$

Proof. Any Riemann-Stieltjes sum for g with respect to f can be expanded and rewritten in the following way:

$$\begin{aligned} & \sum_{k=1}^n g(\xi_k) (f(x_k) - f(x_{k-1})) \\ & = g(\xi_1) (f(x_1) - f(a)) + \dots + g(\xi_n) (f(b) - f(x_{n-1})) \\ & = -f(a)g(a) - \sum_{k=1}^{n-1} f(x_k) (g(\xi_{k+1}) - g(\xi_k)) + f(b)g(b), \end{aligned}$$

where we have introduced $\xi_0 = a$ and $\xi_n = b$. Note that the last sum is a Riemann-Stieltjes sum for f with respect to g on the partition $\{\xi_k\}_{k=0}^n$ with x_k

belonging to $[\xi_k, \xi_{k+1}]$. So this summation identity ensures that if the Riemann-Stieltjes sums for $\int_a^b f(x) dg(x)$ converge as $\|P\| \rightarrow 0$, then the Riemann-Stieltjes sums for $\int_a^b g(x) df(x)$ also converge, and their limit values are related by the desired equation. \square

Example 3.4.2. *Integrating by parts, we get*

$$\int_{-1}^2 x d|x| = x|x| \Big|_{-1}^2 - \int_{-1}^2 |x| dx = 5 \cdot 2^{-1}.$$

Secondly, closely related to the integration by parts formula are the following first and second mean value theorem for Riemann-Stieltjes integrals.

Theorem 3.4.3.

(i) *If $f \in C[a, b]$ and g is increasing on $[a, b]$, then there exists a point $\xi \in [a, b]$ such that*

$$\int_a^b f(x) dg(x) = f(\xi)(g(b) - g(a)).$$

(ii) *If f is monotonic on $[a, b]$ and $g \in C[a, b]$, then there is a point $\xi \in [a, b]$ such that*

$$\int_a^b f(x) dg(x) = f(a) \int_a^\xi dg(x) + f(b) \int_\xi^b dg(x).$$

Proof. (i) Assuming

$$m_f = \min_{x \in [a, b]} f(x) \text{ and } M_f = \max_{x \in [a, b]} f(x),$$

we get by Theorem 3.2.4 (iii) and Theorem 3.3.6

$$m_f \int_a^b dg(x) \leq \int_a^b f(x) dg(x) \leq M_f \int_a^b dg(x).$$

If

$$g(b) - g(a) = \int_a^b dg(x) = 0$$

then g must be constant on $[a, b]$, and hence

$$\int_a^b f(x) dg(x) = 0.$$

In this case, ξ may be chosen to be any number in $[a, b]$. On the other hand, if $g(b) - g(a) > 0$ then by the condition $f \in C[a, b]$ we can find a point $\xi \in [a, b]$ such that

$$f(\xi)(g(b) - g(a)) = \int_a^b f(x) dg(x),$$

thanks to the intermediate value theorem. Thus, the desired result follows.

(ii) Without loss of generality, we may assume that f is increasing. By Theorem 3.4.1, we have

$$\int_a^b f(x) dg(x) = f(b)g(b) - f(a)g(a) - \int_a^b g(x) df(x).$$

An application of (i) to $\int_a^b g(x) df(x)$ produces a point $\xi \in [a, b]$ such that

$$\int_a^b g(x) df(x) = g(\xi)(f(b) - f(a)).$$

Consequently,

$$\begin{aligned} \int_a^b f(x) dg(x) &= f(a)(g(\xi) - g(a)) + f(b)(g(b) - g(\xi)) \\ &= f(a) \int_a^\xi dg(x) + f(b) \int_\xi^b dg(x), \end{aligned}$$

as desired. □

Corollary 3.4.4.

(i) If f is monotonic on $[a, b]$ and $h \in C[a, b]$, then there exists a point $\xi \in [a, b]$ such that

$$\int_a^b f(x)h(x) dx = f(a) \int_a^\xi h(x) dx + f(b) \int_\xi^b h(x) dx.$$

(ii) If f is nonnegative and increasing on $[a, b]$ and $h \in C[a, b]$, then there exists a point $\xi \in [a, b]$ such that

$$\int_a^b f(x)h(x) dx = f(b) \int_\xi^b h(x) dx.$$

(iii) If f is nonnegative and decreasing on $[a, b]$ and $h \in C[a, b]$, then there exists a point $\xi \in [a, b]$ such that

$$\int_a^b f(x)h(x)dx = f(a) \int_a^\xi h(x)dx.$$

Proof. (i) Applying Theorems 3.4.3 (ii) and 3.2.3 to $g(x) = \int_a^x h(t)dt$, we reach the desired assertion.

(ii) and (iii) These can be verified straightforwardly by changing $f(a)$ and $f(b)$ of (i) into 0 respectively. Of course, we have used the fact that this change does not affect the Riemann integral $\int_a^b f(x)h(x)dx$; see also Theorem 2.1.10. \square

As well-known, Corollary 3.4.4 (ii) is Pierre Ossian Bonnet's form of the second mean value theorem for Riemann-Stieltjes integrals.

Thirdly, we conclude this section with a useful result on the evaluation of Riemann-Stieltjes integrals – the change of variable formula.

Theorem 3.4.5. Let $\phi \in C[a, b]$ be monotonic. If $f \in RS_g[\phi(a), \phi(b)]$, then $f \circ \phi \in RS_{g \circ \phi}[a, b]$ and

$$\int_a^b f \circ \phi(x)dg \circ \phi(x) = \int_{\phi(a)}^{\phi(b)} f(t)dg(t).$$

Proof. It is enough to handle the case that ϕ is increasing. Since $\phi \in C[a, b]$, ϕ maps $[a, b]$ onto $[\phi(a), \phi(b)]$. Let $P = \{x_k\}_{k=0}^n$ be an arbitrary partition of $[a, b]$ and choose $\tau_k \in [x_{k-1}, x_k]$ for $1 \leq k \leq n$. Let $\phi(P) = \{\phi(x_k)\}_{k=0}^n$. Then $\phi(P)$ is a partition of $[\phi(a), \phi(b)]$ and

$$\phi(x_{k-1}) \leq \phi(\tau_k) \leq \phi(x_k) \quad \text{for } 1 \leq k \leq n.$$

We have

$$\sum_{k=1}^n f \circ \phi(\tau_k) (g \circ \phi(x_k) - g \circ \phi(x_{k-1})) = \sum_{k=1}^n f(\phi(\tau_k)) (g(\phi(x_k)) - g(\phi(x_{k-1}))).$$

Note that $\phi \in C[a, b]$. So it is uniformly continuous on $[a, b]$. It follows that

$\|\phi(P)\| \rightarrow 0$ as $\|P\| \rightarrow 0$. We conclude

$$\begin{aligned}
 \int_a^b f \circ \phi(x) dg \circ \phi(x) &= \lim_{\|P\| \rightarrow 0} \sum_{k=1}^n f \circ \phi(\tau_k) (g \circ \phi(x_k) - g \circ \phi(x_{k-1})) \\
 &= \lim_{\|P\| \rightarrow 0} \sum_{k=1}^n f(\phi(\tau_k)) (g(\phi(x_k)) - g(\phi(x_{k-1}))) \\
 &= \int_{\phi(a)}^{\phi(b)} f(t) dg(t).
 \end{aligned}$$

□

3.5 Improper Situations

The previously-discussed Riemann-Stieltjes integrals are under the restrictions that both f and g are defined and bounded on a finite interval $[a, b]$. In this section, we will briefly extend the scope of integration by relaxing these restrictions, as we did in the section on the improper Riemann integrals.

Definition 3.5.1.

- (i) Let $f \in RS_g[a, b]$ for every $b \geq a$. Then $\int_a^\infty f(x) dg(x)$ is said to be convergent and equal to $s \in \mathbf{R}$ provided that $s = \lim_{b \rightarrow \infty} \int_a^b f(x) dg(x)$ is finite; otherwise, $\int_a^\infty f(x) dg(x)$ is said to diverge.
- (ii) Let f be defined on $(a, b]$ and $f \in RS_g[c, b]$ for any $c \in (a, b]$. Then $\int_{a^+}^b f(x) dg(x)$ is said to be convergent and equal to $s \in \mathbf{R}$ provided that $s = \lim_{c \rightarrow a^+} \int_c^b f(x) dg(x)$ is finite; otherwise, $\int_{a^+}^b f(x) dg(x)$ is said to diverge.

The improper Riemann-Stieltjes integrals share almost all the properties of the improper Riemann integrals; indeed, the proofs of most results concerning the former are virtually identical to the proofs of the corresponding results about the latter, but with $dg(x)$ replacing dx . Below are some instances.

Theorem 3.5.2.

- (i) Let g be increasing on $[a, \infty)$ and $f \in RS_g[a, b]$ for any $b \geq a$. If $f \geq 0$ on $[a, \infty)$, then $\int_a^\infty f(x) dg(x)$ converges when and only when there exists a constant $\kappa > 0$ such that $\int_a^b f(x) dg(x) \leq \kappa$ for all $b \geq a$.

(ii) Let g be increasing on $(a, b]$ and $f \in RS_g[c, b]$ for any $c \in (a, b]$. If $f \geq 0$ on $(a, b]$, then $\int_{a^+}^b f(x) dg(x)$ converges when and only when there exists a constant $\kappa > 0$ such that $\int_c^b f(x) dg(x) \leq \kappa$ for all $c \geq a$.

Proof. It suffices to prove (i). Clearly $\int_a^b f(x) dg(x)$ is increasing with b and so

$$\lim_{b \rightarrow \infty} \int_a^b f(x) dg(x) = \sup_{b \in [a, \infty)} \int_a^b f(x) dg(x).$$

The assertion follows from the above formula and the observation that $\int_a^\infty f(x) dg(x) \leq \kappa$ whenever the integral converges. \square

Theorem 3.5.3.

(i) Let g be increasing on $[a, \infty)$ and $f \in RS_g[a, b]$ for any $b \geq a$. If $f_2 \geq f_1 \geq 0$ on $[a, \infty)$, then $\int_a^\infty f_1(x) dg(x)$ converges whenever $\int_a^\infty f_2(x) dg(x)$ converges, and hence

$$\int_a^\infty f_1(x) dg(x) \leq \int_a^\infty f_2(x) dg(x).$$

(ii) Let g be increasing on $(a, b]$ and $f \in RS_g[c, b]$ for any $c \in (a, b]$. If $f_2 \geq f_1 \geq 0$ on $(a, b]$, then $\int_{a^+}^b f_1(x) dg(x)$ converges whenever $\int_{a^+}^b f_2(x) dg(x)$ converges, and hence

$$\int_{a^+}^b f_1(x) dg(x) \leq \int_{a^+}^b f_2(x) dg(x).$$

Proof. (i) follows from

$$\int_a^b f_1(x) dg(x) \leq \int_a^b f_2(x) dg(x) \leq \int_a^\infty f_2(x) dg(x).$$

Similarly, we can verify (ii). \square

Theorem 3.5.4.

(i) Let g be increasing on $[a, \infty)$, $f_1, f_2 \in RS_g[a, b]$ for any $b \geq a$ and $f_1, f_2 \geq 0$ on $[a, \infty)$. If $\lim_{x \rightarrow \infty} f_1(x)/f_2(x) = q \in (0, \infty)$, then $\int_a^\infty f_1(x) dg(x)$ and $\int_a^\infty f_2(x) dg(x)$ both converge or both diverge.

(ii) Let g be increasing on $(a, b]$, $f_1, f_2 \in RS_g[c, b]$ for any $c \in (a, b]$ and $f_1, f_2 \geq 0$ on $(a, b]$. If $\lim_{x \rightarrow a^+} f_1(x)/f_2(x) = q \in (0, \infty)$, then $\int_{a^+}^b f_1(x) dg(x)$ and $\int_{a^+}^b f_2(x) dg(x)$ both converge or both diverge.

Proof. To prove (i), we note that there is an $n_0 \in \mathbf{N}$ such that

$$2^{-1}qf_2(x) \leq f_1(x) \leq 2qf_2(x) \text{ for all } x \geq n_0.$$

The conclusion follows by applying Theorem 3.5.2 (i). Of course, the argument for (ii) is similar. \square

Theorem 3.5.5.

(i) Let g be increasing on $[a, \infty)$. If $f \in RS_g[a, b]$ for any $b \geq a$ and $\int_a^\infty |f(x)|dg(x)$ converges, then $\int_a^\infty f(x)dg(x)$ also converges.

(ii) Let g be increasing on $(a, b]$. If $f \in RS_g[c, b]$ for any $c \in (a, b]$ and $\int_{a^+}^b |f(x)|dg(x)$ converges, then $\int_{a^+}^b f(x)dg(x)$ also converges.

Proof. This follows from the inequality $0 \leq |f(x)| - f(x) \leq 2|f(x)|$. \square

As in the case of infinite series, we distinguish between absolutely and conditionally convergent integrals.

Definition 3.5.6. Let g be increasing on $[a, \infty)$ and $f \in RS_g[a, b]$ for every $b \geq a$. Then $\int_a^\infty f(x)dg(x)$ is said to be absolutely convergent provided that $\int_a^\infty |f(x)|dg(x)$ exists. Also, $\int_a^\infty f(x)dg(x)$ is said to be conditionally convergent provided that $\int_a^\infty |f(x)|dg(x)$ does not exist, but $\int_a^\infty f(x)dg(x)$ exists.

For instance, $\int_1^\infty x^{-p} \sin x dx$ is absolutely convergent for $p \in (1, \infty)$. Of course, there are examples of conditional convergence. To see this, we need the following Johann Peter Gustav Lejeune Dirichlet test which is an analog of the Dirichlet test for infinite series.

Theorem 3.5.7. Let f be a nonnegative decreasing function on $[a, \infty)$ with $\lim_{x \rightarrow \infty} f(x) = 0$. If g is bounded on $[a, \infty)$ and $f \in RS_g[a, b]$ for every $b \geq a$, then $\int_a^\infty f(x)dg(x)$ is convergent.

Proof. Integrating by parts, we have

$$\int_a^b f(x)dg(x) = f(b)g(b) - f(a)g(a) + \int_a^b g(x)d(-f(x)).$$

Now since g is bounded, we find $\lim_{b \rightarrow \infty} f(b)g(b) = 0$. Therefore, to complete the argument it suffices to prove that $\int_a^\infty g(x)d(-f(x))$ is convergent. To do so,

we assume $\sup_{x \in [a, \infty)} |g(x)| = M < \infty$ and get that $\int_a^\infty M d(-f(x))$ is convergent due to

$$\lim_{b \rightarrow \infty} \int_a^b M d(-f(x)) = Mf(a).$$

Notice that $-f$ is increasing. So Theorem 3.5.3 (i) yields that $\int_a^\infty |g(x)| d(-f(x))$ is convergent. \square

On the other hand, the following result provides a practical way to determine whether or not an improper integral is conditionally convergent.

Theorem 3.5.8. *Let f be continuous and ϕ decreasing on $[a, \infty)$. If $F(t) = \int_a^t f(x) dx$ is bounded on $[a, \infty)$ and $\lim_{x \rightarrow \infty} \phi(x) = 0$. Then $\int_a^\infty f(x)\phi(x) dx$ converges.*

Proof. This result can be verified by Theorem 3.5.7. Nevertheless, we give a different argument as follows. By assumption, F is bounded on $[a, \infty)$, i.e., $M = \sup_{x \in [a, \infty)} |F(x)| < \infty$. Thus

$$\left| \int_{t_1}^{t_2} f(x) dx \right| = |F(t_2) - F(t_1)| \leq 2M \text{ for } a \leq t_1 < t_2 < \infty.$$

By Corollary 3.4.4 (ii) – Bonnet's form of the second mean value theorem, we have

$$\int_{t_1}^{t_2} f(x)\phi(x) dx = \phi(t_1) \int_{t_1}^{\zeta} f(x) dx$$

for some $\zeta \in [t_1, t_2]$. Hence

$$\left| \int_{t_1}^{t_2} f(x)\phi(x) dx \right| \leq 2M\phi(t_1).$$

By assumption $\lim_{x \rightarrow \infty} \phi(x) = 0$, so that for any $\varepsilon > 0$ there is $t_0 > 0$ such that $t_1 \geq t_0$ implies $\phi(t_1) < \varepsilon$. It follows that for any $\varepsilon > 0$ there is $t_0 > 0$ such that $t_2 > t_1 \geq t_0$ implies

$$\left| \int_{t_1}^{t_2} f(x)\phi(x) dx \right| \leq 2M\varepsilon.$$

From Theorem 2.4.3 (iii) – Cauchy's convergence criterion it follows that $\int_a^\infty f(x)\phi(x) dx$ is convergent. \square

We can obtain many examples of conditional convergence via the following consequence of Theorem 3.5.8.

Corollary 3.5.9. *Let ϕ be a decreasing function on $[a, \infty)$ with $\lim_{x \rightarrow \infty} \phi(x) = 0$. Then $\int_a^\infty \phi(x) \sin x dx$ and $\int_a^\infty \phi(x) \cos x dx$ converge. Moreover, the convergence is absolute or conditional depending on whether $\int_a^\infty \phi(x) dx < \infty$ or $\int_a^\infty \phi(x) dx = \infty$.*

Proof. It is enough to verify the sine-case. Firstly, we have

$$\left| \int_a^t \sin x dx \right| = |\cos a - \cos t| \leq 2.$$

It follows from Theorem 3.5.8 that $\int_a^\infty \phi(x) \sin x dx$ is convergent.

Secondly, suppose that $\int_a^\infty \phi(x) dx$ is finite. We have $|\phi(x) \sin x| \leq \phi(x)$ for $x \in [a, \infty)$. We conclude from the comparison test that $\int_a^\infty |\phi(x) \sin x| dx$ is convergent. It remains to prove that

$$\int_a^\infty \phi(x) dx = \infty \Rightarrow \int_a^\infty |\phi(x) \sin x| dx = \infty.$$

To check this, we show

$$\lim_{n \rightarrow \infty} \int_a^{n\pi} |\phi(x) \sin x| dx = \infty.$$

Let m be a fixed integer $\geq a/\pi$. For $a/\pi \leq m < n$, we have

$$\int_a^{n\pi} |\phi(x) \sin x| dx \geq \int_{m\pi}^{n\pi} |\phi(x) \sin x| dx = \sum_{k=m+1}^n \int_{(k-1)\pi}^{k\pi} |\phi(x) \sin x| dx.$$

Since ϕ is decreasing, if $m+1 \leq k \leq n$, then

$$\begin{cases} \int_{(k-1)\pi}^{k\pi} |\phi(x) \sin x| dx \geq \phi(k\pi) \int_{(k-1)\pi}^{k\pi} |\sin x| dx = 2\phi(k\pi); \\ \int_{k\pi}^{(k+1)\pi} \phi(x) dx \leq \phi(k\pi) \int_{k\pi}^{(k+1)\pi} dx = \pi\phi(k\pi); \\ \int_{(k-1)\pi}^{k\pi} |\phi(x) \sin x| dx \geq \frac{2}{\pi} \int_{k\pi}^{(k+1)\pi} \phi(x) dx. \end{cases}$$

Consequently,

$$\int_a^{n\pi} |\phi(x) \sin x| dx \geq \frac{2}{\pi} \sum_{k=m+1}^n \int_{k\pi}^{(k+1)\pi} \phi(x) dx = \frac{2}{\pi} \int_{(m+1)\pi}^{(n+1)\pi} \phi(x) dx.$$

Letting $n \rightarrow \infty$ in the last estimate, we obtain the desired result:

$$\int_a^\infty |\phi(x) \sin x| dx = \infty.$$

□

Example 3.5.10. In particular, $p \in (0, 1]$ ensures that $\int_1^\infty x^{-p} \sin x dx$ is conditionally convergent since $p \in (0, 1]$ implies $\int_1^\infty x^{-p} dx = \infty$.

Problems

3.1. Find the variations of the following functions on the indicated intervals:

(i) $f(x) = \sin x$ on $[0, 3\pi]$;

(ii) $f(x) = 2x^3 - 3x^2$ on $[-1, 2]$;

(iii) $\{a_k\}_{k=1}^\infty$ is an infinite sequence of real numbers and

$$f(x) = \begin{cases} a_k, & x = (k+1)^{-1} \\ 0, & x \in [0, 1] \setminus \{(k+1)^{-1}\} \end{cases}.$$

3.2. Find $v_f(x)$ for $f(x) = \sin x$ on $[0, 2\pi]$.

3.3. Suppose that $(x = x(t), y = y(t))_{t \in [a, b]}$ represents a continuous arc Γ on the plane. To each partition $P = \{a = t_0, t_1, \dots, t_n = b\}$ there corresponds a sequence of points $\{P_k = (x(t_k), y(t_k))\}_{k=0}^n$. Let $L(P)$ be the sum of the lengths of segments: $\overline{P_0P_1}, \overline{P_1P_2}, \dots$ and $\overline{P_{n-1}P_n}$. Prove that $L = \sup_P L(P) < \infty$ (i.e. Γ is rectifiable) if and only if $x(t)$ and $y(t)$ have bounded variation.

3.4. Let D be a dense subset of the open interval $(a, b) \subseteq \mathbf{R}$. Prove that if two monotonic functions $f_1, f_2 : (a, b) \rightarrow \mathbf{R}$ satisfy the equation $f_1 = f_2$ on D then f_1 and f_2 have the same continuous points.

3.5. Prove that $V_a^b(f_1) < \infty$ and $V_a^b(f_2) < \infty$ imply $V_a^b(f_1 f_2) < \infty$.

3.6. Prove that the following function

$$f(x) = \begin{cases} x \cos \pi(2x)^{-1}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

has no bounded variation on $[0, 1]$.

3.7. For each $k \in \mathbf{N}$ let $f_k \in BV[a, b]$ and $\lim_{k \rightarrow \infty} f_k(x) = f(x)$ where $|f(x)| < \infty$ for each $x \in [a, b]$. Prove that if $\sup_{k \in \mathbf{N}} V_a^b(f_k) < \infty$ then $f \in BV[a, b]$.

3.8. Evaluate the following Riemann-Stieltjes integrals:

(i) $\int_0^4 x^2 d([x])$ where $[x]$ denotes the greatest integer function;

(ii) $\int_0^1 x^3 d(x^2)$;

(iii) $\int_a^c f(x) dg(x)$ where f is continuous and

$$g(x) = \begin{cases} 0, & a \leq x \leq b \\ 1, & b < x \leq c \end{cases}.$$

3.9. Prove Theorem 3.2.4.

3.10. Evaluate $\int_0^{2^{-1}\pi} x d(\sin x)$.

3.11. Suppose $\int_a^b f(x) dg(x)$ exists, $f \in C[a, b]$, and we replace g by a new integrator function h , where $h(x) = g(x)$ if $x \neq c$ for some $c \in (a, b)$. Prove $\int_a^b f(x) dh(x) = \int_a^b f(x) dg(x)$.

3.12. Compare the class of Riemann integrable functions with the class of functions having bounded variation; that is, prove or disprove each of the following statements:

(i) If $f \in R[a, b]$, then $f \in BV[a, b]$;

(ii) If $f \in BV[a, b]$, then $f \in R[a, b]$.

3.13. Construct two functions f and g on $[a, b]$ such that

$$g \in BV[a, b] \text{ and } \lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f(x) dg(x) \neq \int_a^b f(x) dg(x).$$

3.14. Let $g \in BV[a, b]$, $f \in RS_g[a, b]$ and $v_g(x) = V_a^x g$ for each $x \in (a, b]$, prove the following three results.

(i)

$$\left| \int_a^b f(x) dg(x) \right| \leq \int_a^b |f(x)| dv_g(x) \leq \sup_{x \in [a, b]} |f(x)| v_g(b).$$

(ii) If $\{f_n\}_{n=1}^\infty$ is a sequence in $RS_g[a, b]$, then

$$\lim_{n \rightarrow \infty} \sup_{x \in [a, b]} |f_n(x) - f(x)| = 0 \Rightarrow \lim_{n \rightarrow \infty} \int_a^b f_n(x) dg(x) = \int_a^b f(x) dg(x).$$

(iii) If $\{g_n\}_{n=1}^\infty$ is a sequence in $BV[a, b]$ with $\lim_{n \rightarrow \infty} V_a^b(g_n - g) = 0$ then

$$f \in C[a, b] \Rightarrow \lim_{n \rightarrow \infty} \int_a^b f(x) dg_n(x) = \int_a^b f(x) dg(x).$$

3.15. Let g be increasing on $[a, b]$, $f \in RS_g[a, b]$ and $F(x) = \int_a^x f(t) dg(t)$ for $x \in [a, b]$. Prove the following four results.

(i) $F \in BV[a, b]$.

(ii) Every point of continuity of g is also a point of continuity of F .

(iii) If g is increasing on $[a, b]$, then $F'(x)$ exists at each point $x \in (a, b)$ where $g'(x)$ exists and where f is continuous. For such x , one has $F'(x) = f(x)g'(x)$.

(iv) If $h \in C[a, b]$ then $h \in RS_F[a, b]$ and $\int_a^b h(x) dF(x) = \int_a^b h(x)f(x) dg(x)$.

3.16. If $g \in C[0, 1]$ and $\int_a^b f(x) dg(x) = 0$ for all increasing functions f on $[0, 1]$, prove that g must be constant on $[0, 1]$.

3.17. If $s \neq 1$, use the Riemann-Stieltjes integration to derive:

(i) $\sum_{k=1}^n k^{-s} = n^{1-s} + s \int_1^n [x]x^{-(1+s)} dx$;

(ii) $\sum_{k=1}^{2n} (-1)^k k^{-s} = s \int_1^{2n} (2[2^{-1}x] - [x])x^{-(1+s)} dx$.

3.18. Suppose that $f_1, f_2 \in RS_g[a, b]$ and g is increasing on $[a, b]$. Prove the following results.

(i) The Cauchy-Schwarz inequality

$$\left| \int_a^b f_1(x)f_2(x) dg(x) \right| \leq \left(\int_a^b (f_1(x))^2 dg(x) \right)^{\frac{1}{2}} \left(\int_a^b (f_2(x))^2 dg(x) \right)^{\frac{1}{2}}.$$

(ii)

$$\begin{aligned} & 2^{-1} \int_a^b \left(\int_a^b (f_1(y) - f_1(x))(f_2(y) - f_2(x)) dg(y) \right) dg(x) \\ &= (g(b) - g(a)) \int_a^b f_1(x)f_2(x) dg(x) - \left(\int_a^b f_1(x) dg(x) \right) \left(\int_a^b f_2(x) dg(x) \right). \end{aligned}$$

(iii)

$$\left(\int_a^b f_1(x) dg(x) \right) \left(\int_a^b f_2(x) dg(x) \right) \leq (g(b) - g(a)) \int_a^b f_1(x)f_2(x) dg(x)$$

holds for both f_1 and f_2 increasing (or decreasing) on $[a, b]$.

(iv) The reverse inequality in (iii) is valid when f_1 and f_2 are increasing and decreasing on $[a, b]$ respectively.

3.19. *Prove the following three results.*

(i) $\int_1^\infty (x\sqrt{x+1})^{-1} dx$ converges but $\int_0^\infty (x\sqrt{x+1})^{-1} dx$ diverges.

(ii) $\int_{0+}^\infty e^{-x} \sin x^{-1} dx$ converges absolutely;

(iii) $\int_{0+}^1 x^{-1} \cos x^{-1} dx$ converges conditionally.

Chapter 4

Lebesgue-Radon-Stieltjes Integrals

While the Riemann-Stieltjes integral is suitable for many theoretic and applied purposes, it has some technical deficiencies. For instance, this integral no longer commutes with uniform limits - this fact can be verified by the following implication - if

$$f_n(x) = \begin{cases} n^{-1}, & x \in [0, n] \\ 0, & x \in \mathbf{R} \setminus [0, n] \end{cases} \quad \text{for all } n \in \mathbf{N},$$

then f_n converges uniformly to $f = 0$ on \mathbf{R} , and hence

$$\begin{cases} \lim_{n \rightarrow \infty} \sup_{x \in \mathbf{R}} |f_n(x) - f(x)| = 0; \\ \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f_n(x) dx = 1 \neq 0 = \int_{-\infty}^{\infty} \lim_{n \rightarrow \infty} f_n(x) dx. \end{cases}$$

A better route is to abandon the Riemann-Stieltjes integral for the Lebesgue-Radon-Stieltjes integral, named for Henri Leon Lebesgue, Johann Radon and Thomas Joannes Stieltjes, to whom much of the theory of the present topic is due. In this chapter, we try to circumvent the abstract measure-theoretic approach (viewed as the most common expository method) to the Lebesgue-Radon-Stieltjes integrals. Instead, starting with the so-called g -measure of simple sets (g is increasing on \mathbf{R}), we utilize the limit of step functions (for which the Lebesgue-Radon-Stieltjes integral exists) to obtain the integrals for more general functions. Subsequently, we show the basic results and techniques used in Lebesgue-Radon-Stieltjes' integration theory.

4.1 Foundational Material

Let us consider a typical function which is not Riemann integrable.

Example 4.1.1. *If*

$$f(x) = \begin{cases} 1, & x \in \mathbf{Q} \cap (0, 1) \\ 0, & x \notin \mathbf{Q} \cap (0, 1) \end{cases},$$

then $f \notin R[0, 1]$, but there exists a sequence of step functions f_n such that $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for any $x \in [0, 1]$. Thus, the pointwise limit is not closed in $R[0, 1]$.

Indeed, since the rational numbers in $(0, 1)$ can be listed as $\{r_1, r_2, \dots\}$, we can assume

$$h_k(x) = \begin{cases} 1, & x = r_k \\ 0, & x \in [0, 1] \setminus \{r_k\} \end{cases}.$$

It is clear that each h_k is a step function and hence $h_k \in R[0, 1]$. If $f_n = \sum_{k=1}^n h_k$, then it is a step function too and hence belongs to $R[0, 1]$. Note that if x is irrational in $(0, 1)$, 0 or 1, then $f_n(x) = 0 = f(x)$; and if $x \in (0, 1)$ equals some r_{n_0} , then $f_n(r_{n_0}) = 1 = f(r_{n_0})$ when $n \geq n_0$. So $f_n \rightarrow f$ pointwise on $[0, 1]$ as $n \rightarrow \infty$.

Example 4.1.1 motivates the definition of Lebesgue-Radon-Stieltjes' integral. To be more precise, firstly, we need to introduce a way to measure intervals and even simple sets. To do so, let \mathbf{R}^e be the extended real number system consisting of all real numbers, $-\infty$ and $+\infty = \infty$. Sometimes, $[-\infty, \infty]$ is written in place of \mathbf{R}^e . Then \mathbf{R}^e enjoys the following facts:

- (a) $(\pm\infty) + x = x + (\pm\infty) = \pm\infty$ for any $x \in \mathbf{R}$;
- (b) $(\pm\infty) + (\pm\infty) = \pm\infty$;
- (c) $(\pm\infty) \cdot x = x \cdot (\pm\infty) = \pm\infty$ or $\mp\infty$ for any $x > 0$ or $x < 0$;
- (d) $(\pm\infty) \cdot (\pm\infty) = \infty$, $\infty \cdot (-\infty) = (-\infty) \cdot \infty = -\infty$;
- (e) $\infty + (-\infty)$, $(-\infty) + \infty$, $0 \cdot (\pm\infty)$, and $(\pm\infty) \cdot 0$ are undefined.

In addition, by an interval with endpoints $a, b \in \mathbf{R}^e$, we mean that it is one of $[a, b]$, $[a, b)$, $(a, b]$ or (a, b) . Note that $[a, a) = (a, a] = (a, a) = \emptyset$ and $[a, a] = \{a\}$. When $a, b \in \mathbf{R}$, the interval is said to be bounded or finite.

Definition 4.1.2. *Suppose that $g : \mathbf{R} \rightarrow \mathbf{R}$ is an increasing function. For any $c \in \mathbf{R}$ let*

$$g(c^+) = \lim_{0 < x - c \rightarrow 0} g(x) \text{ and } g(c^-) = \lim_{0 < c - x \rightarrow 0} g(x).$$

(i) $m_g(I)$, called the g -measure of an interval I with endpoints $a, b \in \mathbf{R}$, is defined by

$$m_g(I) = \begin{cases} 0, & I = (a, b), a = b \\ g(b^-) - g(a^+), & I = (a, b), a < b \\ g(b^+) - g(a^-), & I = [a, b], a \leq b \\ g(b^+) - g(a^+), & I = (a, b], a \leq b \\ g(b^-) - g(a^-), & I = [a, b], a \leq b \end{cases}.$$

(ii) A set $S \subseteq \mathbf{R}$ is called simple provided $S = \bigcup_{k=1}^n I_k$, the union of a finitely many disjoint finite intervals $\{I_k\}_{k=1}^n$ in \mathbf{R} . In this case, the g -measure of S is defined by $m_g(S) = \sum_{k=1}^n m_g(I_k)$.

Remark 4.1.3. $m_g(S)$ stands for the length of S whenever $g(x) = x$ and S is simple. Furthermore, m_g enjoys three basic properties below:

- (a) $m_g(S) \geq 0$ for any simple set S ;
- (b) $m_g(S_2 \setminus S_1) = m_g(S_2) - m_g(S_1)$ when S_1 and S_2 are simple, $S_1 \subseteq S_2$;
- (c) $m_g(\bigcup_{j=1}^n S_j) = \sum_{j=1}^n m_g(S_j)$ when S_1, S_2, \dots, S_n are pairwise disjoint simple sets.

The following result on the countable additivity of m_g is of fundamental importance in the Lebesgue-Radon-Stieltjes integrals on \mathbf{R} .

Theorem 4.1.4. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be an increasing function. If \mathcal{R} is the collection of the empty set and all simple sets in \mathbf{R} , then m_g satisfies $m_g(\emptyset) = 0$ and $m_g(\bigcup_{j=1}^\infty S_j) = \sum_{j=1}^\infty m_g(S_j)$ for every sequence $\{S_j\}_{j=1}^\infty$ of disjoint sets in \mathcal{R} for which $\bigcup_{j=1}^\infty S_j \in \mathcal{R}$.

Proof. It suffices to check the second property. Suppose that $S = \bigcup_{j=1}^\infty S_j \in \mathcal{R}$. Then

$$m_g(S) = m_g(S \setminus \bigcup_{j=1}^n S_j) + m_g(\bigcup_{j=1}^n S_j) = m_g(S \setminus \bigcup_{j=1}^n S_j) + \sum_{j=1}^n m_g(S_j)$$

thanks to the properties (b) and (c) in Remark 4.1.3. Let $R_n = S \setminus \bigcup_{j=1}^n S_j$. Then $R_{n+1} \subseteq R_n$ for each $n \in \mathbf{N}$ and $\bigcap_{n=1}^\infty R_n = \emptyset$. So the argument will be finished by proving $\lim_{n \rightarrow \infty} m_g(R_n) = 0$. To this end, assume on the contrary that the last limit result is not valid. Then there is a constant $\delta > 0$ such that $m_g(R_n) > \delta$ for each $n \in \mathbf{N}$. But, since g is increasing on \mathbf{R} , for each $n \in \mathbf{N}$ we can find a simple

set I_n and a closed simple set J_n such that $I_n \subseteq J_n \subseteq R_n$ and $m_g(R_n \setminus I_n) < \delta 2^{-n-1}$. Accordingly, it follows from (a)-(b)-(c) in Remark 4.1.3 that

$$m_g(\cap_{j=1}^n I_j) = m_g\left(\cap_{j=1}^n (R_j \setminus (R_j \setminus I_j))\right) \geq m_g(R_n) - \sum_{j=1}^n m_g(R_j \setminus I_j) > 2^{-1}\delta,$$

and consequently, $\{\cap_{j=1}^n J_j\}_{n=1}^\infty$ is a sequence of nonempty, bounded, closed and decreasing sets. Therefore, we use Theorem 1.2.4(i) to obtain

$$\emptyset \neq \cap_{n=1}^\infty J_n \subseteq \cap_{n=1}^\infty R_n = \cap_{n=1}^\infty (S \setminus \cup_{j=1}^n S_j) = S - \cup_{j=1}^\infty S_j = \emptyset,$$

a contradiction. □

Secondly, we modify the Riemann-Stieltjes sums to define summable functions and admissible sequences.

Definition 4.1.5. *Given an increasing function $g : \mathbf{R} \rightarrow \mathbf{R}$.*

(i) *A function $s : \mathbf{R} \rightarrow \mathbf{R}$ is called a step function provided that $\{x \in \mathbf{R} : s(x) \neq 0\}$ is a simple set; that is, there are $n \in \mathbf{N}$, $c_1, \dots, c_n \in \mathbf{R}$ and disjoint finite intervals $\{I_k\}_{k=1}^n$ such that $s = \sum_{k=1}^n c_k 1_{I_k}$. In this case, we write*

$$A_g(s) = \sum_{k=1}^n c_k m_g(I_k)$$

for the g -area of s .

(ii) *A function sequence $\{s_k\}_{k=1}^\infty$ on \mathbf{R} is called admissible for a nonnegative function $f : \mathbf{R} \rightarrow \mathbf{R}^e$ provided that*

- (a) *each s_k is a step function on \mathbf{R} ;*
- (b) *each s_k is nonnegative;*
- (c) *$f \leq \sum_{k=1}^\infty s_k$.*

In this case, we denote by

$$S_g(f) = \inf \left\{ \sum_{k=1}^\infty A_g(s_k) \right\}$$

the g -sum of f , where the infimum ranges over all admissible sequences $\{s_k\}_{k=1}^\infty$ for f .

Remark 4.1.6. A_g is just the Euclidean area under the graph of s when g is the identity. Moreover, there always exists an admissible sequence for any nonnegative $f: \mathbf{R} \rightarrow \mathbf{R}^e$. To this end, we take $I_k = (-k, k)$ and $s_k = 1_{I_k}$. Clearly, $\{s_k\}_{k=1}^\infty$ is admissible for f . This demonstration shows that S_g always exists even if it may be ∞ .

To better understand A_g and S_g , we explore their basic properties as follows.

Theorem 4.1.7. *Let $g: \mathbf{R} \rightarrow \mathbf{R}$ be an increasing function. Then*

(i) $A_g(\cdot)$ has the following four properties:

(a) $A_g(0) = 0$ and $A_g(s) \geq 0$ whenever s is a nonnegative step function on \mathbf{R} ;

(b) $A_g(s_1) \leq A_g(s_2)$ whenever s_1 and s_2 are step functions satisfying $s_1 \leq s_2$ on \mathbf{R} ;

(c) If s is a step function on \mathbf{R} , so are $|s|$, $s^+ = \max\{s, 0\}$, and $s^- = \max\{-s, 0\}$ on \mathbf{R} with

$$A_g(s) = A_g(s^+) - A_g(s^-) \text{ and } A_g(|s|) = A_g(s^+) + A_g(s^-);$$

(d) If $\{s_k\}_{k=1}^n$ are step functions on \mathbf{R} and $\{c_k\}_{k=1}^n$ are finite real numbers, then $s = \sum_{k=1}^n c_k s_k$ is a step function on \mathbf{R} and $A_g(s) = \sum_{k=1}^n c_k A_g(s_k)$.

(ii) $S_g(\cdot)$ has the following four properties:

(a) For $f: \mathbf{R} \rightarrow \mathbf{R}^e$, $S_g(f^\pm) \leq S_g(|f|)$ where $f^\pm = \max\{\pm f, 0\}$;

(b) For $f: \mathbf{R} \rightarrow \mathbf{R}^e$ and $c \in \mathbf{R}$, $S_g(|cf|) = |c|S_g(|f|)$;

(c) For $f_1, f_2: \mathbf{R} \rightarrow \mathbf{R}^e$ with $0 \leq f_1 \leq f_2$, $S_g(f_1) \leq S_g(f_2)$;

(d) If $\{f_k\}_{k=1}^\infty$ is a sequence of functions from \mathbf{R} to \mathbf{R}^e and $\sum_{k=1}^\infty f_k(x)$ converges for any $x \in \mathbf{R}$, then

$$S_g\left(\left|\sum_{k=1}^\infty f_k\right|\right) \leq \sum_{k=1}^\infty S_g(|f_k|).$$

(iii) $A_g(s) = S_g(s)$ holds for any nonnegative step function s on \mathbf{R} .

Proof. It is straightforward to verify (i). Next we prove (ii).

(a) It follows from the definitions of S_g , f^+ and f^- .

(b) If $c = 0$ then the result is trivial. Suppose now $c \neq 0$. Then $\{s_k\}_{k=1}^\infty$ is an admissible sequence for $|f|$ if and only if $\{|c|s_k\}_{k=1}^\infty$ is an admissible sequence for $|cf|$. Moreover,

$$S_g(|cf|) = \inf \left\{ \sum_{k=1}^\infty A_g(|c|s_k) \right\} = |c| \inf \left\{ \sum_{k=1}^\infty A_g(s_k) \right\} = |c|S_g(f).$$

(c) This follows from that if $\{s_k\}_{k=1}^\infty$ is admissible for f_2 then it is admissible for f_1 .

(d) Without loss of generality, we may assume $\sum_{k=1}^\infty S_g(|f_k|) < \infty$: Otherwise, there is nothing to argue. Under this assumption, we have naturally $S_g(|f_k|) < \infty$ for each $k \in \mathbf{N}$. This, along with the definition of the infimum, implies that for any $\varepsilon > 0$ there exists a function sequence $\{s_{k,g}\}_{j=1}^\infty$ which is admissible for $|f_k|$ and satisfies

$$\sum_{j=1}^\infty A_g(s_{k,j}) \leq S_g(|f_k|) + 2^{-k}\varepsilon.$$

Thus

$$\sum_{k,j=1}^\infty A_g(s_{k,j}) = \sum_{k=1}^\infty \left(\sum_{j=1}^\infty A_g(s_{k,j}) \right) \leq \sum_{k=1}^\infty S_g(|f_k|) + \varepsilon.$$

Note that

$$0 \leq \left| \sum_{k=1}^\infty f_k \right| \leq \sum_{k=1}^\infty |f_k| \leq \sum_{k=1}^\infty \left(\sum_{j=1}^\infty s_{k,j} \right) = \sum_{k,j=1}^\infty s_{k,j}.$$

So $\{s_{k,j}\}_{k,j=1}^\infty$ is an admissible sequence for $\left| \sum_{k=1}^\infty f_k \right|$. Further, we get

$$S_g\left(\left| \sum_{k=1}^\infty f_k \right|\right) \leq \sum_{k,j=1}^\infty A_g(s_{k,j}) \leq \sum_{k=1}^\infty S_g(|f_k|) + \varepsilon,$$

and consequently,

$$S_g\left(\left| \sum_{k=1}^\infty f_k \right|\right) \leq \sum_{k=1}^\infty S_g(|f_k|),$$

as desired.

Last of all, we verify (iii). If $s : \mathbf{R} \rightarrow \mathbf{R}$ is a step function, then $S_g(s) \leq A_g(s)$. Assume now that $\{s_k\}_{k=1}^\infty$ is any admissible sequence for s and let $t_n = s - \sum_{j=1}^n s_j$. Then $t_n : \mathbf{R} \rightarrow \mathbf{R}$ is a step function, decreasing and convergent to

$\lim_{n \rightarrow \infty} t_n \leq 0$. Consequently, $t_n^+ = \max\{t_n, 0\}$ decreases and converges to 0. Since

$$K_{1,+} = \{x \in \mathbf{R} : t_1^+(x) > 0\} \text{ and } K_{n,\varepsilon} = \{x \in \mathbf{R} : t_n^+(x) \geq \varepsilon\}$$

are simple sets for any given $\varepsilon > 0$, we conclude

$$\begin{aligned} A_g(t_n^+) &\leq \left(\sup_{x \in K_{n,\varepsilon}} t_n^+(x) \right) m_g(K_{n,\varepsilon}) + \left(\sup_{x \in \mathbf{R} \setminus K_{n,\varepsilon}} t_n^+(x) \right) m_g(\mathbf{R} \setminus K_{n,\varepsilon}) \\ &\leq \left(\sup_{x \in \mathbf{R}} t_1^+(x) \right) m_g(K_{n,\varepsilon}) + \varepsilon m_g(K_{1,+}). \end{aligned}$$

Since

$$K_{n+1,\varepsilon} \subseteq K_{n,\varepsilon} \text{ and } \bigcap_{n=1}^{\infty} K_{n,\varepsilon} = \emptyset,$$

we employ Theorem 4.1.4 to achieve

$$m_g(K_{1,\varepsilon}) = \lim_{n \rightarrow \infty} \sum_{j=1}^n (m_g(K_{j,\varepsilon}) - m_g(K_{j+1,\varepsilon})) = m_g(K_{1,\varepsilon}) - \lim_{n \rightarrow \infty} m_g(K_{n,\varepsilon}),$$

whence $\lim_{n \rightarrow \infty} m_g(K_{n,\varepsilon}) = 0$. The previous discussion implies immediately

$$\lim_{n \rightarrow \infty} \sup_{k \geq n} A_g(t_k^+) \leq \varepsilon m_g(K_{1,+}),$$

which yields $\lim_{n \rightarrow \infty} A_g(t_n^+) = 0$ since $\varepsilon > 0$ is arbitrary. But,

$$s = t_n + \sum_{j=1}^n s_j \leq t_n^+ + \sum_{j=1}^n s_j \Rightarrow A_g(s) \leq A_g(t_n^+) + \sum_{j=1}^n A_g(s_j).$$

Letting $n \rightarrow \infty$ we obtain $A_g(s) \leq \sum_{j=1}^{\infty} A_g(s_j)$ which, together with the definition of $S_g(s)$, yields $A_g(s) \leq S_g(s)$. Accordingly, $A_g(s) = S_g(s)$. \square

Thirdly, the Lebesgue-Radon-Stieltjes integrals are founded on the following facts.

Theorem 4.1.8. *Given an increasing function $g : \mathbf{R} \rightarrow \mathbf{R}$. For $k \in \mathbf{N}$ let $f_k : \mathbf{R} \rightarrow \mathbf{R}$ be a function with $S_g(|f_k|) < \infty$. If $f : \mathbf{R} \rightarrow \mathbf{R}$ satisfies $\lim_{k \rightarrow \infty} S_g(|f - f_k|) = 0$, then:*

(i)

$$S_g(|f|) < \infty \text{ and } S_g(f^\pm) < \infty;$$

(ii)

$$\lim_{k \rightarrow \infty} S_g(|f| - |f_k|) = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} S_g(|f^\pm - f_k^\pm|) = 0;$$

(iii)

$$\lim_{k \rightarrow \infty} S_g(|f_k|) = S_g(|f|) \quad \text{and} \quad \lim_{k \rightarrow \infty} S_g(f_k^\pm) = S_g(f^\pm).$$

Proof. (i) By hypothesis, there is an $n_0 \in \mathbf{N}$ such that $S_g(|f - f_{n_0}|) < \infty$ and $S_g(|f_{n_0}|) < \infty$. So, from Theorem 4.1.7 (ii) (d) it follows that

$$S_g(f^\pm) \leq S_g(|f|) \leq S_g(|f - f_{n_0}|) + S_g(|f_{n_0}|) < \infty.$$

(ii) Note that

$$\max \{ ||f| - |f_k||, |f^+ - f_k^+|, |f^- - f_k^-| \} \leq |f - f_k| \quad \text{for all } k \in \mathbf{N}.$$

So this inequality, along with $\lim_{k \rightarrow \infty} S_g(|f - f_k|) = 0$, implies the desired limits.

(iii) It follows from Theorem 4.1.7 (ii) (d) that for each $k \in \mathbf{N}$,

$$\begin{cases} S_g(|f_k|) \leq S_g(|f|) + S_g(|f_k - f|); \\ S_g(|f|) \leq S_g(|f_k|) + S_g(|f - f_k|); \\ |S_g(|f_k|) - S_g(|f|)| \leq S_g(|f - f_k|). \end{cases}$$

This inequality plus the hypothesis yields $\lim_{k \rightarrow \infty} |S_g(|f_k|) - S_g(|f|)| = 0$. Since $f_k^\pm = f^\pm + (f_k^\pm - f^\pm)$, we can validate $\lim_{k \rightarrow \infty} S_g(f_k^\pm) = S_g(f^\pm)$ in a way similar to showing $\lim_{k \rightarrow \infty} |S_g(|f_k|) - S_g(|f|)| = 0$. \square

Now, we are in a position to introduce the definition of a Lebesgue-Radon-Stieltjes integral.

Definition 4.1.9. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be an increasing function.

(i) For a step function $f : \mathbf{R} \rightarrow \mathbf{R}$, we define the Lebesgue-Radon-Stieltjes integral of f on \mathbf{R} with respect to m_g by

$$\int_{\mathbf{R}} f dm_g = A_g(f^+) - A_g(f^-).$$

(ii) For an arbitrary function $f : \mathbf{R} \rightarrow \mathbf{R}^e$, we say that f is Lebesgue-Radon-Stieltjes integrable on \mathbf{R} with respect to m_g , denoted $f \in LRS_g(\mathbf{R})$ provided that

there exists a sequence $\{s_k\}_{k=1}^{\infty}$ of step functions on \mathbf{R} such that $\lim_{k \rightarrow \infty} S_g(|f - s_k|) = 0$. In this case, we write

$$\int_{\mathbf{R}} f dm_g = S_g(f^+) - S_g(f^-)$$

for the Lebesgue-Radon-Stieltjes integral of f on \mathbf{R} .

We close this section by making two important comments on the above notion.

Remark 4.1.10.

(i) If g in Definition 4.1.9 is the identity on \mathbf{R} , then $\int_{\mathbf{R}} f dm_g$ is called the Lebesgue integral of $f : \mathbf{R} \rightarrow \mathbf{R}^e$ on \mathbf{R} , denoted $\int_{\mathbf{R}} f dm_{id}$. It is clear that if f coincides with the function discussed in Example 4.1.1 on $[0, 1]$ and 0 elsewhere, then f is Lebesgue integrable on \mathbf{R} .

(ii) Definition 4.1.9 produces the following three facts:

- (a) $\int_{\mathbf{R}} 1_I f dm_g = 0$ for any interval $I \subseteq \mathbf{R}$ and g being any constant on I ;
- (b) $\int_{\mathbf{R}} 1_{[c, c]} f dm_g = f(c)(g(c^+) - g(c^-))$ for any $c \in \mathbf{R}$;
- (c) $\int_{\mathbf{R}} 1_I dm_g = m_g(I)$ for any finite interval $I \subset \mathbf{R}$.

4.2 Essential Properties

After introducing the concept of a Lebesgue-Radon-Stieltjes integral, we seek to show its basic properties.

Theorem 4.2.1. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. If $f \in LRS_g(\mathbf{R})$, then $|f|, f^+, f^- \in LRS_g(\mathbf{R})$ with

$$\begin{cases} \int_{\mathbf{R}} f dm_g = \int_{\mathbf{R}} f^+ dm_g - \int_{\mathbf{R}} f^- dm_g; \\ \int_{\mathbf{R}} |f| dm_g = \int_{\mathbf{R}} f^+ dm_g + \int_{\mathbf{R}} f^- dm_g; \\ \left| \int_{\mathbf{R}} f dm_g \right| \leq \int_{\mathbf{R}} |f| dm_g. \end{cases}$$

Proof. Because of $f \in LRS_g(\mathbf{R})$, we can take a sequence $\{s_k\}_{k=1}^{\infty}$ of step functions on \mathbf{R} such that $\lim_{k \rightarrow \infty} S_g(|f - s_k|) = 0$. Since each s_k is a step function \mathbf{R} , so is $|s_k|$ and it follows that $S_g(|s_k|) < \infty$. From Theorem 4.1.8 (ii) it turns out that

$$\lim_{k \rightarrow \infty} S_g(|f| - |s_k|) = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} S_g(|f^{\pm} - s_k^{\pm}|) = 0.$$

By definition, we have $|f|, f^+, f^- \in LRS_g(\mathbf{R})$ with

$$\int_{\mathbf{R}} f dm_g = S_g(f^+) - S_g(f^-) = \int_{\mathbf{R}} f^+ dm_g - \int_{\mathbf{R}} f^- dm_g.$$

Meanwhile, note by Theorem 4.1.7 (i) (c) that

$$A_g(|s_k|) = A_g(s_k^+) + A_g(s_k^-).$$

So it follows that $S_g(|s_k|) = S_g(s_k^+) + S_g(s_k^-)$. This, along with Theorem 4.1.8 (iii), implies

$$\int_{\mathbf{R}} |f| dm_g = S_g(f^+) + S_g(f^-) = \int_{\mathbf{R}} f^+ dm_g + \int_{\mathbf{R}} f^- dm_g,$$

as required.

Finally, the foregoing argument yields the inequality in Theorem 4.2.1 right away. \square

Theorem 4.2.2. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. If*

$$f_k \in LRS_g(\mathbf{R}) \text{ and } \lim_{k \rightarrow \infty} S_g(|f - f_k|) = 0,$$

then $f \in LRS_g(\mathbf{R})$ with

$$\begin{cases} \lim_{k \rightarrow \infty} \int_{\mathbf{R}} f_k dm_g = \int_{\mathbf{R}} f dm_g; \\ \lim_{k \rightarrow \infty} \int_{\mathbf{R}} f_k^\pm dm_g = \int_{\mathbf{R}} f^\pm dm_g; \\ \lim_{k \rightarrow \infty} \int_{\mathbf{R}} |f_k| dm_g = \int_{\mathbf{R}} |f| dm_g. \end{cases}$$

Proof. By Theorem 4.1.7 (ii) (d), $f_k \in LRS_g(\mathbf{R})$ induces a sequence $\{s_{k,g}\}_{j=1}^\infty$ of step functions on \mathbf{R} such that $\lim_{j \rightarrow \infty} S_g(|f_k - s_{k,j}|) = 0$. Thus, for some $j \in \mathbf{N}$ we must have $S_g(|f_k - s_{k,j}|) < k^{-1}$. For any $\varepsilon > 0$, there is an $n_0 \in \mathbf{N}$ such that $S_g(|f - f_k|) < 2^{-1}\varepsilon$ as $k \geq n_0$. Accordingly, if $k \geq 2\varepsilon^{-1}$ then $S_g(|f_k - s_{k,j}|) < 2^{-1}\varepsilon$. Using Theorem 4.1.7 (ii) (d) again, we obtain that

$$k \geq \max\{n_0, 2\varepsilon^{-1}\} \Rightarrow S_g(|f - s_{k,j}|) \leq S_g(|f - f_k|) + S_g(|f_k - s_{k,j}|) < \varepsilon \Rightarrow f \in LRS_g(\mathbf{R}).$$

The limit formulas may be verified by Theorems 4.2.1 and 4.1.8 (iii) plus the definition of the Lebesgue-Radon-Stieltjes integrals. \square

Theorem 4.2.3. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing.*

- (i) *If $0 \leq f \in LRS_g(\mathbf{R})$, then $\int_{\mathbf{R}} f dm_g \geq 0$.*
- (ii) *If $f_1, f_2 \in LRS_g(\mathbf{R})$ satisfy $f_1 \leq f_2$ on \mathbf{R} , then $\int_{\mathbf{R}} f_1 dm_g \leq \int_{\mathbf{R}} f_2 dm_g$.*
- (iii) *If $f_1, \dots, f_n \in LRS_g(\mathbf{R})$ and $c_1, \dots, c_n \in \mathbf{R}$, then*

$$\sum_{k=1}^n c_k f_k \in LRS_g(\mathbf{R}) \text{ with } \int_{\mathbf{R}} \left(\sum_{k=1}^n c_k f_k(x) \right) dm_g(x) = \sum_{k=1}^n c_k \int_{\mathbf{R}} f_k dm_g.$$

Proof. (ii) follows from (i) and (iii), but (i) is derived easily from that $\int_{\mathbf{R}} f dm_g = S_g(f) \geq 0$ as $f \geq 0$. To prove (iii), we may assume $f_k \geq 0$. The condition $f_k \in LRS_g(\mathbf{R})$ produces a sequence $\{s_{k,j}\}_{j=1}^{\infty}$ of step functions on \mathbf{R} such that $\lim_{j \rightarrow \infty} S_g(|f_k - s_{k,j}|) = 0$. Now for each $j \in \mathbf{N}$ we use Theorem 4.1.7 (ii) (d) to get

$$0 \leq S_g \left(\left| \sum_{k=1}^n c_k f_k - \sum_{k=1}^n c_k s_{k,j} \right| \right) \leq \sum_{k=1}^n |c_k| S_g(|f_k - s_{k,j}|).$$

Since n is finite, we conclude that

$$\lim_{j \rightarrow \infty} S_g(|f_k - s_{k,j}|) = 0 \Rightarrow \lim_{j \rightarrow \infty} S_g \left(\left| \sum_{k=1}^n c_k f_k - \sum_{k=1}^n c_k s_{k,j} \right| \right) = 0 \Rightarrow \sum_{k=1}^n c_k f_k \in LRS_g(\mathbf{R}).$$

An application of Theorem 4.2.2 implies

$$\begin{cases} \lim_{j \rightarrow \infty} \int_{\mathbf{R}} s_{k,j} dm_g = \int_{\mathbf{R}} f_k dm_g; \\ \lim_{j \rightarrow \infty} \int_{\mathbf{R}} (\sum_{k=1}^n c_k s_{k,j}) dm_g = \int_{\mathbf{R}} \sum_{k=1}^n c_k f_k dm_g. \end{cases}$$

By virtue of the foregoing limits and the following equalities

$$\int_{\mathbf{R}} \left(\sum_{k=1}^n c_k s_{k,j} \right) dm_g = A_g \left(\sum_{k=1}^n c_k s_{k,j} \right) = \sum_{k=1}^n c_k \int_{\mathbf{R}} s_{k,j} dm_g,$$

we obtain the desired result. \square

The following result, a consequence of Theorem 4.2.3, will be very useful in evaluating some Lebesgue-Radon-Stieltjes integrals.

Corollary 4.2.4. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing and $I \subseteq \mathbf{R}$ a union of finitely many disjoint finite intervals $\{I_k\}_{k=1}^n$. Then $1_I f \in LRS_g(\mathbf{R})$ if and only if $1_{I_k} f \in LRS_g(\mathbf{R})$ for any $k = 1, \dots, n$. In this case,*

$$\int_{\mathbf{R}} 1_I f dm_g = \sum_{k=1}^n \int_{\mathbf{R}} 1_{I_k} f dm_g.$$

Proof. If $1_I f \in LRS_g(\mathbf{R})$, then there is a sequence of step functions $\{s_k\}_{k=1}^\infty$ on \mathbf{R} such that $\lim_{k \rightarrow \infty} S_g(|1_I f - s_k|) = 0$. So, $\lim_{k \rightarrow \infty} S_g(|1_{I_j} f - 1_{I_j} s_k|) = 0$ and $1_{I_j} f \in LRS_g(\mathbf{R})$ for each $j = 1, 2, \dots, n$. Note that $1_I f = \sum_{k=1}^n 1_{I_k} f$ on \mathbf{R} . Then by Theorem 4.2.3 (iii) we get

$$\int_{\mathbf{R}} 1_I f dm_g = \int_{\mathbf{R}} \sum_{k=1}^n 1_{I_k} f dm_g = \sum_{k=1}^n \int_{\mathbf{R}} 1_{I_k} f dm_g.$$

Conversely, if $1_{I_k} f \in LRS_g(\mathbf{R})$ for each $k = 1, 2, \dots, n$, then there is a sequence of step functions $\{s_{k,j}\}_{j=1}^\infty$ on \mathbf{R} such that $\lim_{j \rightarrow \infty} S_g(|1_{I_k} f - s_{k,j}|) = 0$. This yields $\lim_{j \rightarrow \infty} S_g(|f 1_I - \sum_{k=1}^n s_{k,j}|) = 0$. Since $\{\sum_{k=1}^n s_{k,j}\}_{j=1}^\infty$ is also a sequence of step functions on \mathbf{R} , $f 1_I \in LRS_g(\mathbf{R})$ follows. \square

We give an example to demonstrate how Corollary 4.2.4 is applied.

Example 4.2.5. Let

$$g(x) = \begin{cases} 0, & x < 2 \\ 1, & x \geq 2 \end{cases}.$$

Then an application of Corollary 4.2.4 and Remark 4.1.10 (ii) gives

$$\begin{aligned} & \int_{\mathbf{R}} x^6 1_{[2,4]}(x) dm_g(x) \\ &= \int_{\mathbf{R}} x^6 1_{[2,2]}(x) dm_g(x) + \int_{\mathbf{R}} x^6 1_{(2,4)}(x) dm_g(x) + \int_{\mathbf{R}} x^6 1_{[4,4]}(x) dm_g(x) \\ &= 2^6 (g(2^+) - g(2^-)) + 0 + 4^6 (g(4^+) - g(4^-)) \\ &= 2^6. \end{aligned}$$

Motivated by Theorem 4.2.3 (i), we consider the so-called null functions and sets with respect to m_g .

Definition 4.2.6. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing.

(i) $f : \mathbf{R} \rightarrow \mathbf{R}^e$ is called an m_g -null function provided that $f \in LRS_g(\mathbf{R})$ and $\int_{\mathbf{R}} |f| dm_g = 0$.

(ii) $E \subseteq \mathbf{R}$ is called an m_g -null set provided that $1_E \in LRS_g(\mathbf{R})$ and $\int_{\mathbf{R}} 1_E dm_g = 0$.

(iii) A property \mathcal{P} on a function $f : \mathbf{R} \rightarrow \mathbf{R}^e$ is said to hold almost everywhere with respect to m_g (denoted m_g -a.e.) on \mathbf{R} provided that \mathcal{P} holds everywhere in \mathbf{R} except possibly on some m_g -null subset of \mathbf{R} .

Clearly, f is an m_g -null function if and only if $S_g(|f|) = 0 = S_g(f)$. Furthermore, the following result not only gives the visible description of an m_g -null function, but also shows that changing the values of a function on an m_g -null set does not affect the integral of the function – two functions that are equal m_g -a.e. can be regarded as identical in the context of integration theory.

Theorem 4.2.7. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing.*

- (i) $f : \mathbf{R} \rightarrow \mathbf{R}^e$ is an m_g -null function if and only if $f = 0$ m_g -a.e. on \mathbf{R} .
 (ii) If $f_1, f_2 : \mathbf{R} \rightarrow \mathbf{R}^e$ satisfy $f_1 = f_2$ m_g -a.e. on \mathbf{R} , and $f_1 \in LRS_g(\mathbf{R})$, then

$$f_2 \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} f_1 dm_g = \int_{\mathbf{R}} f_2 dm_g.$$

Proof. (i) Suppose that $f : \mathbf{R} \rightarrow \mathbf{R}^e$ is a m_g -null function. Then $S_g(|f|) = 0$. For each $n - 1 \in \mathbf{N}$ let

$$E_1 = \{x \in \mathbf{R} : |f(x)| \geq 1\} \text{ and } E_n = \{x \in \mathbf{R} : n^{-1} \leq |f(x)| < (n-1)^{-1}\}.$$

It is evident that $0 \leq 1_{E_n} \leq n|f|$ for $n \in \mathbf{N}$ and so that

$$0 \leq S_g(1_{E_n}) \leq S_g(n|f|) = nS_g(|f|) = 0.$$

Let $E = \{x \in \mathbf{R} : f(x) \neq 0\}$. Then for any $x \in E$ there is a unique $n_0 \in \mathbf{N}$ such that $x \in E_{n_0}$ and hence $1_{E_n}(x) = 1$ whenever $n = n_0$ and $1_{E_n}(x) = 0$ whenever $n \neq n_0$. This means that

$$1_E(x) = 1 = \sum_{n=1}^{\infty} 1_{E_n}(x) \text{ for all } x \in E.$$

However, if $x \notin E$, then $x \notin E_n$ for all $n \in \mathbf{N}$, and hence $1_E(x) = 0 = \sum_{n=1}^{\infty} 1_{E_n}(x)$. Therefore, $1_E = \sum_{n=1}^{\infty} 1_{E_n}$. This, together with Theorem 4.1.7 (ii) (d), implies

$$0 \leq S_g(1_E) \leq \sum_{n=1}^{\infty} S_g(1_{E_n}) = 0,$$

and so 1_E is an m_g -null function; namely, E is an m_g -null set; i.e., $f = 0$ m_g -a.e. on \mathbf{R} .

Conversely, suppose $f = 0$ m_g -a.e. on \mathbf{R} . Then $E = \{x \in \mathbf{R} : f(x) \neq 0\}$ is an m_g -null set. For $n \in \mathbf{N}$ let

$$F_n = \{x \in \mathbf{R} : n-1 < |f(x)| \leq n\}.$$

Naturally, each F_n is an m_g -null set. Set $f_n = n1_{F_n}$. Then, to any $x \in E$ there corresponds a unique $n_0 \in \mathbf{N}$ such that $x \in F_{n_0}$, and so

$$f_n(x) = \begin{cases} n_0, & n = n_0 \\ 0, & n \neq n_0 \end{cases}.$$

Consequently,

$$\sum_{n=1}^{\infty} f_n(x) = n_0 \geq |f(x)| \text{ for all } x \in E.$$

Of course, if $x \in \mathbf{R} \setminus E$, then $x \notin F_n$ for all $n \in \mathbf{N}$ and hence $\sum_{n=1}^{\infty} f_n(x) = 0 = f(x)$. Thus, $|f| \leq \sum_{n=1}^{\infty} f_n$ on \mathbf{R} . Using Theorem 4.1.7 (ii) (c)-(d), we find

$$0 \leq S_g(|f|) \leq \sum_{n=1}^{\infty} S_g(f_n) = \sum_{n=1}^{\infty} nS_g(1_{F_n}) = 0,$$

which means that f is an m_g -null function.

(ii) Since $f_1 = f_2$ m_g -a.e. on \mathbf{R} is equivalent to $f_1 - f_2 = 0$ m_g -a.e. on \mathbf{R} , the desired result follows from the foregoing (i) and Theorem 4.2.3 (iii). \square

4.3 Convergence Theorems

We will present the dominated convergence theorem and its consequent monotone convergence theorem as well as Pierre Fatou's lemma for the Lebesgue-Radon-Stieltjes integrals. As an application of the monotone convergence theorem, we will prove that the Lebesgue-Radon-Stieltjes integrals actually generalize the Riemann-Stieltjes integrals.

Here is the dominated convergence theorem.

Theorem 4.3.1. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. Suppose that $\{f_k\}_{k=1}^{\infty}$ is a sequence in $LRS_g(\mathbf{R})$ such that $|f_k| \leq h \in LRS_g(\mathbf{R})$. If $\lim_{k \rightarrow \infty} f_k = f$ m_g -a.e. on \mathbf{R} , then*

$$f \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} \lim_{k \rightarrow \infty} f_k dm_g = \int_{\mathbf{R}} f dm_g = \lim_{k \rightarrow \infty} \int_{\mathbf{R}} f_k dm_g.$$

Proof. To prove the conclusion, we need two auxiliary facts.

Fact 1. If $\{s_k\}_{k=1}^{\infty}$ is a decreasing and nonnegative sequence of step functions on \mathbf{R} and $\lim_{k \rightarrow \infty} s_k = 0$ m_g -a.e. on \mathbf{R} , then $\lim_{k \rightarrow \infty} \int_{\mathbf{R}} s_k dm_g = 0$.

Indeed, $\lim_{k \rightarrow \infty} s_k(x) = s(x) \geq 0$ exists for all $x \in \mathbf{R}$ due to the above-given condition. Note that $\lim_{k \rightarrow \infty} s_k = 0$ m_g -a.e. on \mathbf{R} . So, s is an m_g -null function. Furthermore, from Definitions 4.1.5 and 4.1.9, and Theorem 4.1.7 (d)'s it follows that

$$\begin{aligned}
 0 &\leq \int_{\mathbf{R}} (s_k - s) dm_g \\
 &= S_g(s_k - s) \\
 &= S_g\left(\sum_{n=k+1}^{\infty} (s_{n-1} - s_n)\right) \\
 &\leq \sum_{n=k+1}^{\infty} (S_g(s_{n-1}) - S_g(s_n)) \\
 &= S_g(s_k) - \lim_{m \rightarrow \infty} S_g(s_m) \\
 &\rightarrow 0 \text{ as } k \rightarrow \infty.
 \end{aligned}$$

This implies $\lim_{k \rightarrow \infty} \int_{\mathbf{R}} s_k dm_g = 0$.

Fact 2. If each t_n is the limit of an increasing and nonnegative sequence of step functions on \mathbf{R} , if $S_g(t_1) < \infty$, and if $\{t_n\}_{n=1}^{\infty}$ is a decreasing sequence and tends to 0 m_g -a.e. on \mathbf{R} , then $\lim_{n \rightarrow \infty} S_g(t_n) = 0$.

To check *Fact 2*, let $\epsilon > 0$ and $\epsilon_m = \epsilon 2^{-m}$ for $m \in \mathbf{N}$. Using the assumptions on $\{t_n\}_{n=1}^{\infty}$ and *Fact 1*, we can get step functions u_m on \mathbf{R} such that $S_g(t_m - u_m) < \epsilon_m$, $0 \leq u_m \leq t_m$. Then $v_n = \inf_{1 \leq m \leq n} u_m$ is a step function on \mathbf{R} , $\{v_n\}_{n=1}^{\infty}$ is decreasing and since $0 \leq v_n \leq u_n \leq t_n$, it follows that $\lim_{n \rightarrow \infty} v_n = 0$ m_g -a.e. on \mathbf{R} , and thus, by *Fact 1*, we get

$$\lim_{n \rightarrow \infty} S_g(v_n) = \lim_{n \rightarrow \infty} \int_{\mathbf{R}} v_n dm_g = 0.$$

However, $\{t_n\}_{n=1}^{\infty}$ is a decreasing sequence and enjoys

$$t_n = \inf_{1 \leq m \leq n} t_m = \inf_{1 \leq m \leq n} (u_m + (t_m - u_m)) \leq v_n + \sum_{m=1}^n (t_m - u_m),$$

which derives that there is an $n_0 \in \mathbf{N}$ such that

$$S_g(t_n) \leq S_g(v_n) + \sum_{m=1}^n S_g(t_m - u_m) \leq 2\epsilon \text{ as } n > n_0.$$

Thus, the result follows.

Now, let us prove the desired convergence result. Let $\varepsilon > 0$ and $\varepsilon_k = 2^{-k}\varepsilon$ for $k \in \mathbf{N}$. To each k there corresponds a step function s_k with $S_g(|f_k - s_k|) \leq \varepsilon_k$. Then

$$\sup_{k \geq n} |f_k - s_k| \leq \sum_{k=n}^{\infty} |f_k - s_k|$$

and hence

$$\begin{aligned} S_g\left(\limsup_{m \rightarrow \infty} \sup_{k \geq m} |f_k - s_k|\right) &\leq S_g\left(\sup_{k \geq n} |f_k - s_k|\right) \\ &\leq \sum_{k=n}^{\infty} S_g(f_k - s_k) \\ &\leq 2^{1-n}\varepsilon \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

This last estimation tells us that $\lim_{n \rightarrow \infty} \sup_{k \geq n} |f_k - s_k|$ is an m_g -null function but also equals 0 m_g -a.e. on \mathbf{R} . Note that $\lim_{k \rightarrow \infty} f_k = f$ m_g -a.e. on \mathbf{R} . So it follows that $\lim_{k \rightarrow \infty} s_k = f$ m_g -a.e. on \mathbf{R} . Of course, this gives that

$$t_n = \sup_{k, m \geq n} |s_k - s_m| \Rightarrow \lim_{n \rightarrow \infty} t_n = \lim_{n \rightarrow \infty} \sup_{k, m \geq n} |s_k - s_m| = 0 \text{ } m_g\text{-a.e. on } \mathbf{R}.$$

In addition, $\{t_n\}_{n=1}^{\infty}$ is a decreasing sequence with the following property

$$t_1 \leq 2 \sup_{k \in \mathbf{N}} |s_k| \leq 2 \sup_{k \in \mathbf{N}} |f_k| + 2 \sup_{k \in \mathbf{N}} |f_k - s_k| \leq 2h + 2 \sum_{k=1}^{\infty} |f_k - s_k|,$$

which gives

$$S_g(t_1) \leq 2S_g(h) + 2 \sum_{k=1}^{\infty} S_g(|f_k - s_k|) \leq 2S_g(h) + 2\varepsilon.$$

In other words, t_n satisfies the conditions of *Fact 2*. So, $\lim_{n \rightarrow \infty} S_g(t_n) = 0$. Nevertheless, since

$$|f - s_n| = \lim_{k \rightarrow \infty} |s_k - s_n| \leq t_n \text{ } m_g\text{-a.e. on } \mathbf{R},$$

it follows that

$$\lim_{k \rightarrow \infty} S_g(|f - s_k|) = 0 \text{ and so } f \in LRS_g(\mathbf{R}).$$

Finally, $S_g(|f_k - s_k|) \leq \varepsilon_k$, together with Theorems 4.1.7 (ii) (d) and 4.1.8, implies

$$\lim_{k \rightarrow \infty} \int_{\mathbf{R}} f_k dm_g = \int_{\mathbf{R}} f dm_g.$$

□

As a direct consequence of the dominated convergence theorem, we get the following monotone convergence theorem.

Theorem 4.3.2. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. Suppose that $\{f_k\}_{k=1}^{\infty}$ is an increasing sequence in $LRS_g(\mathbf{R})$ enjoying $\lim_{k \rightarrow \infty} f_k = f$ m_g -a.e. on \mathbf{R} . Then $f \in LRS_g(\mathbf{R})$ if and only if $\lim_{k \rightarrow \infty} \int_{\mathbf{R}} f_k dm_g$ exists. In this case,*

$$\lim_{k \rightarrow \infty} \int_{\mathbf{R}} f_k dm_g = \int_{\mathbf{R}} f dm_g.$$

Proof. If $f \in LRS_g(\mathbf{R})$, then $\int_{\mathbf{R}} f dm_g$ is finite, and hence the estimate $|f_k - f_1| \leq |f - f_1|$ and Theorem 4.3.1 give

$$\lim_{k \rightarrow \infty} \int_{\mathbf{R}} f_k dm_g = \int_{\mathbf{R}} f dm_g.$$

On the other hand, if $\lim_{k \rightarrow \infty} \int_{\mathbf{R}} f_k dm_g$ is finite, then by Theorems 4.1.7 (ii) (d) and 4.2.3 (iii) we have

$$\begin{aligned} S_g(f - f_1) &= S_g\left(\sum_{k=2}^{\infty} (f_k - f_{k-1})\right) \\ &\leq \sum_{k=2}^{\infty} S_g(f_k - f_{k-1}) \\ &= \lim_{n \rightarrow \infty} \sum_{k=2}^n \int_{\mathbf{R}} (f_k - f_{k-1}) dm_g \\ &= \lim_{n \rightarrow \infty} \sum_{k=2}^n \left(\int_{\mathbf{R}} f_k dm_g - \int_{\mathbf{R}} f_{k-1} dm_g \right) \\ &= - \int_{\mathbf{R}} f_1 dm_g + \lim_{n \rightarrow \infty} \int_{\mathbf{R}} f_n dm_g \\ &< \infty. \end{aligned}$$

Consequently, there exists a sequence of nonnegative step functions $\{s_j\}_{j=1}^{\infty}$ on \mathbf{R} such that

$$\begin{cases} f - f_1 \leq h = \sum_{j=1}^{\infty} s_j = \lim_{n \rightarrow \infty} \sum_{j=1}^n s_j; \\ S_g(f - f_1) \leq \sum_{j=1}^{\infty} A_g(s_j) < \infty. \end{cases}$$

With the last result, we have

$$S_g\left(h - \sum_{j=1}^n s_j\right) \leq \sum_{j=n+1}^{\infty} S_g(s_j) \leq \sum_{j=n+1}^{\infty} A_g(s_j) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

This, plus Theorem 4.2.2, implies $h \in LRS_g(\mathbf{R})$. Since

$$f_1 \in LRS_g(\mathbf{R}) \text{ and } \lim_{k \rightarrow \infty} (f_k - f_1) = f - f_1 \text{ } m_g - a.e. \text{ on } \mathbf{R},$$

we conclude from Theorem 4.3.1 that

$$f \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} (f - f_1) dm_g = \lim_{k \rightarrow \infty} \int_{\mathbf{R}} (f_k - f_1) dm_g,$$

as desired. \square

Given a sequence $\{a_n\}_{n=1}^{\infty}$ in \mathbf{R} bounded from below. If $b_n = \inf\{a_m : m \geq n\}$, then $\{b_n\}_{n=1}^{\infty}$ is an increasing sequence of real numbers, and hence $\lim_{n \rightarrow \infty} b_n \in \mathbf{R}^e$. Using this notion, we have the following Fatou's lemma.

Lemma 4.3.3. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. If $\{f_k\}_{k=1}^{\infty}$ is a nonnegative function sequence in $LRS_g(\mathbf{R})$ such that $\lim_{k \rightarrow \infty} f_k = f$ m_g -a.e. on \mathbf{R} and $\lim_{n \rightarrow \infty} \inf_{m \geq n} \int_{\mathbf{R}} f_m dm_g$ is finite, then*

$$f \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} f dm_g \leq \liminf_{n \rightarrow \infty} \inf_{m \geq n} \int_{\mathbf{R}} f_m dm_g.$$

Proof. For each $k \in \mathbf{N}$ let $h_k = \inf_{j \geq k} f_j$. Then $\{h_k\}_{k=1}^{\infty}$ is increasing and satisfied with

$$f = \lim_{k \rightarrow \infty} h_k = \sup_{k \in \mathbf{N}} h_k \text{ } m_g - a.e. \text{ on } \mathbf{R}.$$

By Theorem 4.3.2 and the fact that $0 \leq h_k \leq f_j$ for each $j \geq k$, we derive

$$\int_{\mathbf{R}} f dm_g = \lim_{k \rightarrow \infty} \int_{\mathbf{R}} h_k dm_g \leq \lim_{k \rightarrow \infty} \inf_{j \geq k} \int_{\mathbf{R}} f_j dm_g,$$

as desired. \square

In the rest of this section, we use the above-discussed convergence results to show that the Lebesgue-Radon-Stieltjes integral is a generalization of the Riemann-Stieltjes integral.

Theorem 4.3.4. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing and continuous, and suppose that $f : \mathbf{R} \rightarrow \mathbf{R}$ equals 0 outside the finite interval $[a, b]$ in \mathbf{R} . If $f \in RS_g[a, b]$, then*

$$f \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} f dm_g = \int_a^b f(x) dg(x).$$

Proof. Assume $f \in RS_g[a, b]$. For each $k \in \mathbf{N}$, partition $[a, b]$ into k subintervals with equal length k^{-1} . Denote this partition by $P = \{x_k\}_{k=0}^n$. Define

$$G_k = \sum_{j=1}^k m_j 1_{[x_{j-1}, x_j)} \text{ and } H_k = \sum_{j=1}^k M_j 1_{[x_{j-1}, x_j)},$$

where

$$m_j = \inf\{f(x) : x \in [x_{j-1}, x_j]\} \text{ and } M_j = \sup\{f(x) : x \in [x_{j-1}, x_j]\}.$$

Clearly, $\{G_k\}_{k=1}^\infty$ and $\{H_k\}_{k=1}^\infty$ are increasing and decreasing sequences respectively, but also G_k and H_k belong to $LRS_g(\mathbf{R})$ thanks to both that g is continuous on \mathbf{R} and that $f \in RS_g[a, b]$ implies

$$\begin{aligned} \lim_{k \rightarrow \infty} \int_{\mathbf{R}} G_k dm_g &= \lim_{k \rightarrow \infty} A_g(G_k) \\ &= \lim_{k \rightarrow \infty} \sum_{j=1}^k m_j (g(x_j) - g(x_{j-1})) \\ &= \int_a^b f(x) dg(x) \end{aligned}$$

and

$$\begin{aligned} \lim_{k \rightarrow \infty} \int_{\mathbf{R}} H_k dm_g &= \lim_{k \rightarrow \infty} A_g(H_k) \\ &= \lim_{k \rightarrow \infty} \sum_{j=1}^k M_j (g(x_j) - g(x_{j-1})) \\ &= \int_a^b f(x) dg(x). \end{aligned}$$

Putting $G = \lim_{k \rightarrow \infty} G_k$ and $H = \lim_{k \rightarrow \infty} H_k$, we employ Theorem 4.3.2 to deduce $G, H \in LRS_g(\mathbf{R})$ for which $G \leq f \leq H$ m_g -a.e. on \mathbf{R} . Notice also that

$$H_k - G_k \geq 0 \text{ and } \lim_{k \rightarrow \infty} (H_k - G_k) = H - G \text{ } m_g\text{-a.e. on } \mathbf{R}.$$

So, from Theorems 4.2.7 and 4.3.2 it follows that

$$\begin{aligned}
 0 &\leq \int_{\mathbf{R}} (H - G) dm_g \\
 &= \lim_{k \rightarrow \infty} \int_{\mathbf{R}} (H_k - G_k) dm_g \\
 &= \lim_{k \rightarrow \infty} \int_{\mathbf{R}} H_k dm_g - \lim_{k \rightarrow \infty} \int_{\mathbf{R}} G_k dm_g \\
 &= 0.
 \end{aligned}$$

The above estimates indicate that $H - G = 0$ and so $G = f = H$ m_g -a.e. on \mathbf{R} . Accordingly, we obtain

$$f \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} f dm_g = \lim_{k \rightarrow \infty} \int_{\mathbf{R}} G_k dm_g = \int_a^b f(x) dg(x).$$

□

Theorem 4.3.4 presents a practical way to evaluate many Lebesgue-Radon-Stieltjes integrals.

Example 4.3.5. *Let*

$$g(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & x \geq 0 \end{cases}.$$

Then

$$\begin{aligned}
 &\int_{\mathbf{R}} 1_{[-1,1]}(x) e^x dm_g(x) \\
 &= \int_{\mathbf{R}} 1_{[-1,-1]}(x) e^x dm_g(x) + \int_{\mathbf{R}} 1_{(-1,0)}(x) e^x dm_g(x) + \int_{\mathbf{R}} 1_{[0,1]}(x) e^x dm_g(x) \\
 &= e^{-1} (g((-1)^+) - g((-1)^-)) + 0 + \int_0^1 e^x dg(x) \\
 &= \int_0^1 e^x e^{-x} dx \\
 &= 1.
 \end{aligned}$$

It is worth mentioning that the generalization breaks down whenever the Riemann-Stieltjes integral is improper. For instance, when $p \in (0, 1]$, $\int_1^\infty x^{-p} \sin x dx$ is convergent – see also Example 3.5.10, but $\int_1^\infty x^{-p} |\sin x| dx = \infty$ and so $\int_{[1,\infty)} x^{-p} \sin x dx$ does not exist. However, the following assertion is valid.

Corollary 4.3.6. *Let $f, g : \mathbf{R} \rightarrow \mathbf{R}$ be two functions such that g is increasing and continuous and that f belongs to $RS_g[-n, n]$ for any $n \in \mathbf{N}$ with $\lim_{n \rightarrow \infty} \int_{-n}^n |f(x)| dg(x) < \infty$. Then*

$$f \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} f dm_g = \lim_{n \rightarrow \infty} \int_{-n}^n f(x) dg(x).$$

Proof. Since $\{|f1_{[-n,n]}|\}_{n=1}^{\infty}$ is an increasing sequence and convergent to $|f|$ but also both $|f1_{[-n,n]}|$ and $f1_{[-n,n]}$ lie in $LRS_g(\mathbf{R})$ due to Theorem 4.3.4, a further application of Theorem 4.3.4 yields

$$\lim_{n \rightarrow \infty} \int_{\mathbf{R}} |f1_{[-n,n]}| dm_g = \lim_{n \rightarrow \infty} \int_{-n}^n |f| dg < \infty.$$

This limit result and Theorem 4.3.2 derive $|f| \in LRS_g(\mathbf{R})$. Consequently, the pointwise convergence $\lim_{n \rightarrow \infty} f1_{[-n,n]} = f$, via Theorem 4.3.1, gives

$$f \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} f dm_g = \lim_{n \rightarrow \infty} \int_{-n}^n f(x) dg(x).$$

□

4.4 Extension via Measurability

In the sequel, we deal with the Lebesgue-Radon-Stieltjes integration on any subset of the real line in terms of measurable functions.

Definition 4.4.1. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. A function $f : \mathbf{R} \rightarrow \mathbf{R}^e$ is called m_g -measurable provided that there is a sequence $\{s_k\}_{k=1}^{\infty}$ of step functions on \mathbf{R} such that $\lim_{k \rightarrow \infty} s_k = f$ m_g -a.e. on \mathbf{R} .*

A good but important thing is that integrability can be characterized in terms of measurability.

Theorem 4.4.2. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. Then $f \in LRS_g(\mathbf{R})$ if and only if f is m_g -measurable and $S_g(|f|) < \infty$.*

Proof. The necessity follows from the proof of Theorem 4.3.1 with $f_k = f$ for all $k \in \mathbf{N}$. To argue the sufficiency, suppose that $\{s_k\}_{k=1}^{\infty}$ is a sequence of step functions on \mathbf{R} with $\lim_{k \rightarrow \infty} s_k = f$ m_g -a.e. on \mathbf{R} , and let $S_g(|f|) < \infty$. Using the

second part of the argument for Theorem 4.3.2 we can obtain $h \in LRS_g(\mathbf{R})$ such that $|f| \leq h$. If

$$t_k = \begin{cases} -h, & s_k < -h \\ s_k, & |s_k| \leq h \\ h, & s_k > h \end{cases},$$

then $|t_k| \leq h$ and $t_k \in LRS_g(\mathbf{R})$ by Theorem 4.2.3 (iii). It is easy to see that $\lim_{k \rightarrow \infty} t_k = f$ m_g -a.e. on \mathbf{R} . So Theorem 4.3.1 is used to deduce $f \in LRS_g(\mathbf{R})$. \square

Any step function on \mathbf{R} is obviously m_g -measurable and so $1 = \lim_{n \rightarrow \infty} 1_{(-n, n)}$ is m_g -measurable. Moreover, the m_g -measurability is preserved under the usual analytic operations - more precisely - we have the following property.

Theorem 4.4.3. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing.*

- (i) *If $f : \mathbf{R} \rightarrow \mathbf{R}^e$ is m_g -measurable, so are $|f|$, f^+ and f^- .*
- (ii) *If $f_1, f_2 : \mathbf{R} \rightarrow \mathbf{R}^e$ are m_g -measurable, so are $f_1 \pm f_2$, $f_1 f_2$, f_1 / f_2 (where $f_2 \neq 0$ m_g -a.e. on \mathbf{R}), $\max\{f_1, f_2\}$, and $\min\{f_1, f_2\}$.*
- (iii) *If $\{f_j\}_{j=1}^\infty$ is a sequence of m_g -measurable functions from \mathbf{R} to \mathbf{R}^e , so are $\limsup f_j$ and $\liminf f_j$. In particular, if the m_g -measurable sequence $\{f_j\}_{j=1}^\infty$ converges to f m_g -a.e. on \mathbf{R} , then f is m_g -measurable.*

Proof. (i) This follows from the definition of an m_g -measurable function on \mathbf{R} .

(ii) It suffices to verify that $1/f$ is m_g -measurable when $f : \mathbf{R} \rightarrow \mathbf{R}^e$ is m_g -measurable and $f \neq 0$ m_g -a.e. on \mathbf{R} . Assuming the last when-statement, we get a sequence of step functions $\{s_k\}_{k=1}^\infty$ on \mathbf{R} such that $\lim_{k \rightarrow \infty} s_k = f$ m_g -a.e. on \mathbf{R} . If

$$t_k(x) = \begin{cases} (s_k(x))^{-1}, & s_k(x) \neq 0 \\ 1, & s_k(x) = 0, x \in (-k, k) \\ 0, & s_k(x) = 0, x \in \mathbf{R} \setminus (-k, k) \end{cases},$$

then $\{t_k\}_{k=1}^\infty$ is a sequence of step functions on \mathbf{R} and convergent to $1/f$ m_g -a.e. on \mathbf{R} , and hence $1/f$ is m_g -measurable.

(iii) Noting the following formulas:

$$\begin{cases} \limsup f_j = \lim_{j \rightarrow \infty} \sup_{k \geq j} f_k = \lim_{j \rightarrow \infty} \lim_{l \rightarrow \infty} \sup_{l \geq k \geq j} f_k; \\ \liminf f_j = \lim_{j \rightarrow \infty} \inf_{k \geq j} f_k = \lim_{j \rightarrow \infty} \lim_{l \rightarrow \infty} \inf_{l \geq k \geq j} f_k, \end{cases}$$

we just prove the special case that the m_g -a.e. limit function f of the m_g -measurable function sequence $\{f_j\}_{j=1}^\infty$ is m_g -measurable. To do so, let $\mathcal{S} = \{S_k\}_{k=1}^\infty$ be a sequence of disjoint simple sets such that

$$\mathbf{R} = \bigcup_{k=1}^\infty S_k \text{ and } 0 < m_g(S_k) < \infty.$$

Then the positive function $h : \mathbf{R} \rightarrow \mathbf{R}^e$, defined via putting $h(x) = (k^2 m_g(S_k))^{-1}$ for $x \in S_k$, belongs to $LRS_g(\mathbf{R})$ thanks to Theorem 4.3.2. Since f_j is m_g -measurable, by Definition 4.4.1 there exists a sequence of step functions $\{s_{j,k}\}_{k=1}^\infty$ on \mathbf{R} such that $s_{j,k}$ equals 0 outside a finite union of sets from \mathcal{S} , and $\lim_{k \rightarrow \infty} s_{j,k} = f_j$ m_g -a.e. on \mathbf{R} . Accordingly, if

$$q_{j,k} = h s_{j,k} (h + |s_{j,k}|)^{-1} \text{ and } q_j = \begin{cases} h f_j (h + |f_j|)^{-1}, & f_j \neq \pm\infty \\ \pm h, & f_j = \pm\infty \end{cases},$$

then $\lim_{k \rightarrow \infty} q_{j,k} = q_j$ m_g -a.e. on \mathbf{R} ; $q_{j,k}$ is a step function on \mathbf{R} and hence q_j is m_g -measurable; moreover $|q_j| \leq h$ and $\{q_j\}_{j=1}^\infty$ converges m_g -a.e. on \mathbf{R} to the following function

$$q = \begin{cases} h f (h + |f|)^{-1}, & f \neq \pm\infty \\ \pm h, & f = \pm\infty \end{cases}.$$

From $h \in LRS_g(\mathbf{R})$, Theorems 4.3.1 and Theorem 4.4.2 it follows that q is m_g -measurable, and one can define a sequence of step functions $\{s_j\}_{j=1}^\infty$, each being 0 outside a finite union of sets from \mathcal{S} , so that $|s_j| < h$ and $\lim_{j \rightarrow \infty} s_j = q$ m_g -a.e. on \mathbf{R} . Accordingly, $(h - |s_j|)^{-1} s_j h$ is a step function on \mathbf{R} and convergent to f (see also the definition of q) m_g -a.e. on \mathbf{R} , and consequently, f is m_g -measurable. □

In the light of Theorem 4.4.2, we introduce the following concept.

Definition 4.4.4. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing.

(i) A subset E of \mathbf{R} is called the Lebesgue-Radon-Stieltjes g -measurable or m_g -measurable provided that 1_E is m_g -measurable.

(ii) If $E \subseteq \mathbf{R}$ is m_g -measurable, then $m_g(E) = \int_{\mathbf{R}} 1_E dm_g = S_g(1_E)$ is said to be the Lebesgue-Radon-Stieltjes g -measure of E .

(iii) Let $E \subseteq \mathbf{R}$ and $f : \mathbf{R} \rightarrow \mathbf{R}^e$. We say $f \in LRS_g(E)$ provided that $f1_E \in LRS_g(\mathbf{R})$ and E is m_g -measurable. In this case, the Lebesgue-Radon-Stieltjes integral of f on E is defined by $\int_E f dm_g = \int_{\mathbf{R}} f1_E dm_g$.

In particular, if $g = id$, i.e., $g(x) = x$ for all $x \in \mathbf{R}$, in (i)-(ii)-(iii), then the corresponding terms are the Lebesgue measurable, the Lebesgue measure, the Lebesgue integral.

Remark 4.4.5. A few words on Definition 4.4.4 are arranged below:

(a) It is possible to have $m_g(E) = \infty$ – this means that E has infinite g -measure. Of course, if $m_g(E) < \infty$ then E is said to have finite g -measure – in particular - $m_g(E) = m_g(S) < \infty$ whenever $E = S$ is a simple set.

(b) When $f \in LRS_g(\mathbf{R})$ and E is m_g -measurable, (a) and Theorem 4.4.2 are applied to imply that $f1_E$ is m_g -measurable with $S_g(|f1_E|) \leq S_g(|f|) < \infty$, and hence $f1_E \in LRS_g(\mathbf{R})$ by Theorem 4.4.2.

(c) Given an m_g -measurable interval I in \mathbf{R} , the theorems about $LRS_g(\mathbf{R})$ given above have analogues for $LRS_g(I)$, obtained by multiplying all functions involved by 1_I . In order to avoid unnecessary repeating, we will not state such theorems explicitly, yet will use them freely.

Theorem 4.4.6. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing and \mathcal{M}_g comprise all m_g -measurable subsets of \mathbf{R} . Then

(i) \mathcal{M}_g is a σ -ring in the sense of:

$$\begin{cases} \emptyset, \mathbf{R} \in \mathcal{M}_g; \\ E \in \mathcal{M}_g \Rightarrow E^c = \mathbf{R} \setminus E \in \mathcal{M}_g; \\ E_j \in \mathcal{M}_g \Rightarrow \bigcup_{j=1}^{\infty} E_j \in \mathcal{M}_g. \end{cases}$$

(ii) m_g is a measure on \mathcal{M}_g in the sense of:

$$\begin{cases} m_g(\emptyset) = 0; \\ m_g(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} m_g(E_j) \text{ for any sequence of disjoint sets } \{E_j\}_{j=1}^{\infty} \subseteq \mathcal{M}_g. \end{cases}$$

(iii) m_g is not only increasing - $m_g(E_1) \leq m_g(E_2)$ as $E_1 \subseteq E_2$ in \mathcal{M}_g , but also countably subadditive -

$$m_g(\bigcup_{j=1}^{\infty} E_j) \leq \sum_{j=1}^{\infty} m_g(E_j) \text{ for any sequence } \{E_j\}_{j=1}^{\infty} \subseteq \mathcal{M}_g.$$

(iv) $m_g(\mathbf{R}) < \infty$ if and only if m_g is bounded on all simple sets if and only if $\lim_{n \rightarrow \infty} g(\pm n) < \infty$.

(v) If $f \in LRS_g(\mathbf{R})$, then

$$L_g(E) = \int_E f dm_g \text{ and } L_g^\pm(E) = \int_E f^\pm dm_g \text{ for any } E \in \mathcal{M}_g$$

are countably additive in the sense of:

$$L_g(\cup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} L_g(E_j) \text{ and } L_g^\pm(\cup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} L_g^\pm(E_j)$$

for any sequence of disjoint sets $\{E_j\}_{j=1}^{\infty} \subseteq \mathcal{M}_g$.

(vi) $\int_E f dm_g = 0$ for all $E \in \mathcal{M}_g$ if and only if f is an m_g -null function.

(vii) A function $f : \mathbf{R} \rightarrow \mathbf{R}^e$ is m_g -measurable if and only if $E(f; \lambda) = \{x \in \mathbf{R} : f(x) > \lambda\}$ is m_g -measurable for any $\lambda \in \mathbf{R}$.

Proof. (i) Below is a simple observation. If $f \geq 0$ is m_g -measurable and $E, E_1, E_2 \in \mathcal{M}_g$, then $1_E f$ is m_g -measurable by Theorem 4.4.3 (ii), and the following two formulas

$$1_{E^c \setminus \mathbf{R}} f = (1 - 1_E) f \text{ and } 1_{E_1 \cup E_2} f = \max\{1_{E_1} f, 1_{E_2} f\}$$

are valid. The preceding observation, along with Definition 4.4.4 and Theorems 4.4.2-4.4.3, establishes the desired result.

(ii) Trivially, we have $m_g(\emptyset) = 0$. Suppose that $\{E_j\}_{j=1}^{\infty}$ is a sequence of disjoint sets in \mathcal{M}_g . Then it is easy to see that $1_{\cup_{j=1}^{\infty} E_j}$ is the pointwise limit of the increasing sequence $\{\sum_{j=1}^n 1_{E_j}\}_{n=1}^{\infty}$ and so that

$$\begin{aligned} m_g(\cup_{j=1}^{\infty} E_j) &= \int_{\mathbf{R}} 1_{\cup_{j=1}^{\infty} E_j} dm_g \\ &= \lim_{n \rightarrow \infty} \int_{\mathbf{R}} 1_{\cup_{j=1}^n E_j} dm_g \\ &= \lim_{n \rightarrow \infty} \sum_{j=1}^n \int_{\mathbf{R}} 1_{E_j} dm_g \\ &= \sum_{j=1}^{\infty} m_g(E_j). \end{aligned}$$

(iii) On the one hand, if $E_1 \subseteq E_2$ in \mathcal{M}_g , then (ii) gives

$$m_g(E_2) = m_g(E_1 \cup (E_2 \setminus E_1)) = m_g(E_1) + m_g(E_2 \setminus E_1) \geq m_g(E_1).$$

On the other hand, for any sequence $\{E_j\}_{j=1}^\infty$ of sets in \mathcal{M}_g let

$$F_1 = E_1 \text{ and } F_j = E_j \cap \left(\bigcup_{k=1}^{j-1} E_k\right)^c \text{ for } j > 1.$$

Then

$$\begin{cases} F_j \cap F_l = \emptyset \text{ for } j \neq l; \\ E_j \subseteq F_j \in \mathcal{M}_g; \\ \bigcup_{j=1}^\infty E_j = \bigcup_{k=1}^\infty F_k \in \mathcal{M}_g. \end{cases}$$

Accordingly, (ii) yields

$$m_g\left(\bigcup_{j=1}^\infty E_j\right) = m_g\left(\bigcup_{k=1}^\infty F_k\right) = \sum_{k=1}^\infty m_g(F_k) \leq \sum_{k=1}^\infty m_g(E_k).$$

(iv) This follows from $\mathbf{R} = \bigcup_{j=1}^\infty (-j, j)$ and (ii) as well as

$$\begin{aligned} m_g(\mathbf{R}) &= m_g((-1, 1)) + \sum_{j=2}^\infty m_g((-j, j) \setminus (-j+1, j-1)) \\ &= m_g((-1, 1)) + \lim_{n \rightarrow \infty} \sum_{j=2}^n \left(m_g((-j, j)) - m_g((-j+1, j-1)) \right) \\ &= \lim_{n \rightarrow \infty} m_g((-n, n)). \end{aligned}$$

(v) Since $f \in LRS_g(\mathbf{R})$, the result follows from Theorems 4.2.3 (iii) and 4.3.1 for $L_g^+(\cdot)$ as well as for $L_g^-(\cdot)$ and $L_g(\cdot)$.

(vi) The sufficiency is evident. The necessity follows from the fact that $L_g(E) = 0$ implies $L_g^\pm(E) = 0$ as $E \in \mathcal{M}_g$.

(vii) Assume that the statement after the if and only if is valid. Then

$$E(f; \infty) = \{x \in \mathbf{R} : f(x) = \infty\} = \bigcap_{j=1}^\infty E(f; j)$$

and

$$E(f; -\infty) = \{x \in \mathbf{R} : f(x) = -\infty\} = \bigcap_{j=1}^\infty (E(f; -j))^c$$

are m_g -measurable. Moreover,

$$\begin{aligned} E(f; j, k) &= \{x \in \mathbf{R} : jk^{-1} < f(x) \leq (j+1)k^{-1}\} \\ &= E(f; jk^{-1}) \cap (E(f; (j+1)k^{-1}))^c \end{aligned}$$

is m_g -measurable for $k \in \mathbf{N}$ and $j \in \mathbf{Z}$. Consequently, the characteristic functions $1_{E(f; \pm\infty)}$ and $1_{E(f; j, k)}$ are m_g -measurable. This gives that

$$f_k = k^{-1} \sum_{j=-\infty}^{\infty} j 1_{E(f; j, k)} + k 1_{E(f; \infty)} - k 1_{E(f; -\infty)}$$

is m_g -measurable and so is $f = \lim_{k \rightarrow \infty} f_k$ m_g -a.e. on \mathbf{R} .

Conversely, fix $\lambda \in \mathbf{R}$ and suppose that f is m_g -measurable. Then there exists a sequence of step functions $\{s_k\}_{k=1}^{\infty}$ which converges to f m_g -a.e. on \mathbf{R} . Note that

$$\begin{aligned} \{x \in \mathbf{R} : \liminf s_k(x) > \lambda\} &= \bigcup_{j=1}^{\infty} \liminf_{k \rightarrow \infty} \{x \in \mathbf{R} : s_k(x) \geq \lambda + j^{-1}\} \\ &= \bigcup_{j=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{x \in \mathbf{R} : s_m(x) \geq \lambda + j^{-1}\} \end{aligned}$$

is m_g -measurable since each $\{x \in \mathbf{R} : s_m(x) \geq \lambda + j^{-1}\}$ is a simple set. So, it follows that $E(f; \lambda)$ is m_g -measurable since $f - \liminf_{k \rightarrow \infty} s_k$ is an m_g -null function. \square

4.5 Double, Iterated and Generic Integrals

Motivated by Theorems 4.2.3(iii) and 4.3.1 or 4.3.2 as well as Remark 3.3.5 or Theorem 3.3.4, in this section we settle the question of exchanging the order of the integration in a double integral, and of evaluating an integral as an iterated one, but also employ the established results to introduce an integration over an abstract set.

Definition 4.5.1. Let $g_1, g_2 : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. Then

- (i) A rectangle in $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R}$ is a set of the form $I_1 \times I_2$ where I_1, I_2 are finite intervals in \mathbf{R} . A simple set in \mathbf{R}^2 is a union of finitely many disjoint rectangles.
- (ii) The $g_1 \otimes g_2$ -measure of a rectangle $I_1 \times I_2$ in \mathbf{R}^2 is determined by

$$m_{g_1 \otimes g_2}(I_1 \times I_2) = m_{g_1}(I_1) m_{g_2}(I_2).$$

The $g_1 \otimes g_2$ -measure of a simple set $S = \cup_{k=1}^n (I_{1,k} \times I_{2,k})$ in \mathbf{R}^2 is defined by

$$m_{g_1 \otimes g_2}(S) = \sum_{k=1}^n m_{g_1 \otimes g_2}(I_{1,k} \times I_{2,k}) = \sum_{k=1}^n m_{g_1}(I_{1,k}) m_{g_2}(I_{2,k}).$$

(iii) A function $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ is said to be a step function on \mathbf{R}^2 provided that $\{x = (x_1, x_2) \in \mathbf{R}^2 : f(x) \neq 0\}$ is a simple set $S = \cup_{k=1}^n I_{1,k} \times I_{2,k}$ for which f equals a constant c_k on $I_{1,k} \times I_{2,k}$.

(iv) The Lebesgue-Radon-Stieltjes double integral on \mathbf{R}^2 of a function $f : \mathbf{R}^2 \rightarrow \mathbf{R}^e$, denoted

$$\int_{\mathbf{R}^2} f dm_{g_1 \otimes g_2} = \int_{\mathbf{R}^2} f^+ dm_{g_1 \otimes g_2} - \int_{\mathbf{R}^2} f^- dm_{g_1 \otimes g_2},$$

is defined by a process similar to that used for the one-variable Lebesgue-Radon-Stieltjes integral in the Section 4.1. The class of all functions f with $\int_{\mathbf{R}^2} f dm_{g_1} \otimes m_{g_2}$ being finite is written as $LRS_{g_1 \otimes g_2}(\mathbf{R}^2)$.

Under this definition, we can analogously establish the two-dimensional concepts and results corresponding to those presented in Sections 4.1-4.4. For instance, a function $f : \mathbf{R}^2 \rightarrow \mathbf{R}^e$ is $m_{g_1 \otimes g_2}$ -measurable provided that there is a sequence of step functions $\{s_k\}_{k=1}^\infty$ on \mathbf{R}^2 such that $\lim_{k \rightarrow \infty} s_k = f$ $m_{g_1 \otimes g_2}$ -a.e. on \mathbf{R}^2 . When evaluating a Lebesgue-Radon-Stieltjes double integral, we are naturally led to a problem how to convert the integral to iterated integrals. This problem can be handled via Fubini-Tonelli's theorem (named in honor of Guido Fubini and Leonida Tonelli) as follows.

Theorem 4.5.2. For two increasing functions $g_1, g_2 : \mathbf{R} \rightarrow \mathbf{R}$ let $f : \mathbf{R}^2 \rightarrow \mathbf{R}^e$ be $m_{g_1 \otimes g_2}$ -measurable.

(i) $f(\cdot, x_2)$ and $f(x_1, \cdot)$ are m_{g_1} -measurable for m_{g_2} -almost all $x_2 \in \mathbf{R}$ and m_{g_2} -measurable for m_{g_1} -almost all $x_1 \in \mathbf{R}$ respectively.

(ii) The following three conditions are equivalent:

(a) $\int_{\mathbf{R}} \left(\int_{\mathbf{R}} |f| dm_{g_1} \right) dm_{g_2} < \infty;$

(b) $\int_{\mathbf{R}^2} |f| dm_{g_1 \otimes g_2} < \infty;$

(c) $\int_{\mathbf{R}} \left(\int_{\mathbf{R}} |f| dm_{g_2} \right) dm_{g_1} < \infty.$

Under any one of (a)-(b)-(c), the following three integrals

$$\int_{\mathbf{R}} \left(\int_{\mathbf{R}} f dm_{g_1} \right) dm_{g_2}; \quad \int_{\mathbf{R}^2} f dm_{g_1 \otimes g_2}; \quad \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f dm_{g_2} \right) dm_{g_1}$$

are finite and equal.

Proof. First of all, if f is a step function on \mathbf{R}^2 , then there are real numbers $\{c_k\}_{k=1}^n$ and disjoint rectangles $\{I_{1,k} \times I_{2,k}\}_{k=1}^n$ such that

$$f = \sum_{k=1}^n c_k 1_{I_{1,k} \times I_{2,k}}.$$

Accordingly, (i) follows right away, and (ii) follows from the related definitions and properties of the integrals via

$$\begin{aligned} \int_{\mathbf{R}^2} f dm_{g_1 \otimes g_2} &= \sum_{k=1}^n c_k \int_{\mathbf{R}^2} 1_{I_{1,k} \times I_{2,k}} dm_{g_1 \otimes g_2} \\ &= \sum_{k=1}^n c_k m_{g_1}(I_{1,k}) m_{g_2}(I_{2,k}) \\ &= \sum_{k=1}^n c_k \int_{\mathbf{R}} \left(\int_{\mathbf{R}} 1_{I_{1,k}} dm_{g_1} \right) 1_{I_{2,k}} dm_{g_2} \\ &= \sum_{k=1}^n c_k \int_{\mathbf{R}} \left(\int_{\mathbf{R}} 1_{I_{2,k}} dm_{g_2} \right) 1_{I_{1,k}} dm_{g_1} \\ &= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f dm_{g_2} \right) dm_{g_1} \\ &= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f dm_{g_1} \right) dm_{g_2}. \end{aligned}$$

Next, we consider the general case. Since $f = f^+ - f^-$ and $|f| = f^+ + f^-$, without loss of generality we may assume $f \geq 0$.

(i) Note that $f = \lim_{n \rightarrow \infty} 1_{Q_n} \min\{f, n\}$ holds $m_{g_1 \otimes g_2}$ -a.e. on \mathbf{R}^2 where Q_n is the rectangle $\{(x_1, x_2) \in \mathbf{R}^2 : |x_1|, |x_2| \leq n\}$ for $n \in \mathbf{N}$. If $f_n = 1_{Q_n} \min\{f, n\}$, then $0 \leq f_n \leq n$ on Q_n and $f = 0$ on $\mathbf{R}^2 \setminus Q_n$. Since f is $m_{g_1 \otimes g_2}$ -measurable, so is f_n . Therefore, it suffices to verify (i) for each f_n .

By definition there are a sequence of step functions $\{s_{n,k}\}_{k=1}^\infty$ on \mathbf{R}^2 and a set $E_n \subseteq Q_n$ such that

$$\begin{cases} m_{g_1 \otimes g_2}(E_n) = 0; \\ 0 \leq s_{n,k} \leq n; \\ \lim_{k \rightarrow \infty} s_{n,k}(x) = f_n(x) \text{ for } x \in Q_n \setminus E_n. \end{cases}$$

Since E_n is an $m_{g_1 \otimes g_2}$ -null set, for any $\varepsilon > 0$ there are a sequence of rectangles

$\{R_{n,j}\}_{j=1}^{\infty}$ and a sequence of numbers $\{c_{n,j}\}_{j=1}^{\infty}$ such that

$$1_{E_n} \leq \sum_{j=1}^{\infty} c_{n,j} 1_{R_{n,j}} \quad \text{and} \quad \sum_{j=1}^{\infty} c_{n,j} m_{g_1 \otimes g_2}(R_{n,j}) < \varepsilon.$$

Accordingly,

$$\begin{aligned} \int_{\mathbf{R}} \left(\int_{\mathbf{R}} 1_{E_n} dm_{g_2} \right) dm_{g_1} &\leq \sum_{j=1}^{\infty} \int_{\mathbf{R}} \left(\int_{\mathbf{R}} c_{n,j} 1_{R_{n,j}} dm_{g_2} \right) dm_{g_1} \\ &= \sum_{j=1}^{\infty} \int_{R_{n,j}} c_{n,j} 1_{R_{n,j}} dm_{g_1 \otimes g_2} \\ &< \varepsilon \end{aligned}$$

and

$$\begin{aligned} \int_{\mathbf{R}} \left(\int_{\mathbf{R}} 1_{E_n} dm_{g_1} \right) dm_{g_2} &\leq \sum_{j=1}^{\infty} \int_{\mathbf{R}} \left(\int_{\mathbf{R}} c_{n,j} 1_{R_{n,j}} dm_{g_1} \right) dm_{g_2} \\ &= \sum_{j=1}^{\infty} \int_{R_j} c_{n,j} 1_{R_{n,j}} dm_{g_1 \otimes g_2} \\ &< \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, it follows that

$$\int_{\mathbf{R}} 1_{E_n}(x_1, \cdot) dm_{g_2} = 0 \quad \text{for } m_{g_1} - \text{almost all } x_1 \in \mathbf{R}$$

and

$$\int_{\mathbf{R}} 1_{E_n}(\cdot, x_2) dm_{g_1} = 0 \quad \text{for } m_{g_2} - \text{almost all } x_2 \in \mathbf{R}.$$

In other words, $1_{E_n}(x_1, \cdot)$ and $1_{E_n}(\cdot, x_2)$ are m_{g_2} -null and m_{g_1} -null functions for almost all x_1 and x_2 in \mathbf{R} respectively, and consequently,

$$\lim_{j \rightarrow \infty} s_{n,j}(x_1, \cdot) = f_n(x_1, \cdot) \quad \text{for } m_{g_1} - \text{almost all } x_1 \in \mathbf{R}$$

and

$$\lim_{j \rightarrow \infty} s_{n,j}(\cdot, x_2) = f_n(\cdot, x_2) \quad \text{for } m_{g_2} - \text{almost all } x_2 \in \mathbf{R}.$$

Accordingly, (i) is valid for f_n and then for f .

(ii) It is enough to show that

$$\int_{\mathbf{R}} \left(\int_{\mathbf{R}} f dm_{g_1} \right) dm_{g_2} < \infty \Leftrightarrow \int_{\mathbf{R}^2} f dm_{g_1 \otimes g_2} < \infty.$$

To this end, let $f_n = 1_{Q_n} \min\{f, n\}$ be as above. Then

$$\int_{\mathbf{R}^2} f_n dm_{g_1 \otimes g_2} = \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f_n dm_{g_1} \right) dm_{g_2}$$

which follows from the dominated convergence theorem and the validity of this formula for the step functions $\{s_{n,j}\}_{j=1}^{\infty}$ above.

Let $\int_{\mathbf{R}} \left(\int_{\mathbf{R}} f dm_{g_1} \right) dm_{g_2} < \infty$. Since $f_n = \min\{f, n\} 1_{Q_n}$ belongs to $LRS_{g_1 \otimes g_2}(\mathbf{R}^2)$ for each $n \in \mathbf{N}$ and converges to f as $n \rightarrow \infty$, we use the last formula for f_n to obtain

$$\int_{\mathbf{R}^2} f_n dm_{g_1 \otimes g_2} \leq \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f dm_{g_1} \right) dm_{g_2} < \infty,$$

which ensures $f \in LRS_{g_1 \otimes g_2}(\mathbf{R}^2)$.

Conversely, if $f \in LRS_{g_1 \otimes g_2}(\mathbf{R}^2)$, then a combination of both the last formula for f_n and the monotone convergence theorem gives

$$\lim_{n \rightarrow \infty} \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f_n dm_{g_1} \right) dm_{g_2} = \lim_{n \rightarrow \infty} \int_{\mathbf{R}^2} f_n dm_{g_1 \otimes g_2} = \int_{\mathbf{R}^2} f dm_{g_1 \otimes g_2} < \infty.$$

Since $\{f_n\}_{n=1}^{\infty}$ increases to f , for each fixed $x_2 \in \mathbf{R}$ the sequence $\{f_n(\cdot, x_2)\}_{n=1}^{\infty}$ increases to $f(\cdot, x_2)$, and consequently, if

$$F_n(x_2) = \int_{\mathbf{R}} f_n(\cdot, x_2) dm_{g_1} \quad \text{and} \quad F(x_2) = \int_{\mathbf{R}} f(\cdot, x_2) dm_{g_1},$$

then $\{F_n(x_2)\}_{n=1}^{\infty}$ increases to $F(x_2)$. Since each F_n is m_{g_2} -measurable by the argument for (i), so is F . This fact and the monotone convergence theorem yield

$$\lim_{n \rightarrow \infty} \int_{\mathbf{R}} F_n dm_{g_2} = \int_{\mathbf{R}} F dm_{g_2}.$$

The previous argument actually implies

$$\int_{\mathbf{R}} \left(\int_{\mathbf{R}} f dm_{g_1} \right) dm_{g_2} = \int_{\mathbf{R}^2} f dm_{g_1 \otimes g_2} < \infty.$$

□

Example 4.5.3. Let $g_1(x) = g_2(x) = x$ on \mathbf{R} and

$$f(x_1, x_2) = \begin{cases} x_1 \exp(-x_1^2(1+x_2^2)), & (x_1, x_2) \in [0, \infty) \times [0, \infty) \\ 0, & (x_1, x_2) \in \mathbf{R}^2 \setminus [0, \infty) \times [0, \infty) \end{cases}.$$

Then

$$\begin{aligned} \int_{\mathbf{R}} \left(\int_{\mathbf{R}} f \, dm_{g_1} \right) dm_{g_2} &= \int_0^\infty \left(\int_0^\infty x_1 \exp(-x_1^2(1+x_2^2)) \, dx_1 \right) dx_2 \\ &= \int_0^\infty 2^{-1}(1+x_2^2)^{-1} \, dx_2 \\ &= 2^{-2}\pi \end{aligned}$$

and hence by Theorem 4.5.2 (ii) and the substitution $z_1 = x_1 x_2$ we get

$$\begin{aligned} 2^{-2}\pi &= \int_0^\infty \left(\int_0^\infty x_1 \exp(-x_1^2(1+x_2^2)) \, dx_2 \right) dx_1 \\ &= \int_0^\infty \exp(-x_1^2) \left(\int_0^\infty \exp(-z_1^2) \, dz_1 \right) dx_1 \\ &= \left(\int_0^\infty \exp(-t^2) \, dt \right)^2. \end{aligned}$$

The Fubini-Tonelli theorem can be used to establish the well-known Hermann Minkowski's inequality.. To this end, we need the following Hölder inequality which is named after Otto Hölder and viewed as a fundamental inequality in analysis.

Theorem 4.5.4. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing, $1 < p < \infty$ and $q = p(p-1)^{-1}$. If $f_1, f_2 : \mathbf{R} \rightarrow \mathbf{R}^e$ are m_g -measurable, then

$$\int_{\mathbf{R}} |f_1 f_2| \, dm_g \leq \left(\int_{\mathbf{R}} |f_1|^p \, dm_g \right)^{\frac{1}{p}} \left(\int_{\mathbf{R}} |f_2|^q \, dm_g \right)^{\frac{1}{q}}.$$

Proof. Two cases are considered below.

Case 1:

$$\int_{\mathbf{R}} |f_1|^p \, dm_g = \infty \quad \text{or} \quad \int_{\mathbf{R}} |f_2|^q \, dm_g = \infty.$$

The desired inequality is trivial.

Case 2:

$$0 < \int_{\mathbf{R}} |f_1|^p dm_g < \infty \text{ and } 0 < \int_{\mathbf{R}} |f_2|^q dm_g < \infty.$$

For this, we consider $f(t) = p^{-1} + q^{-1}t^q - t$ for $t \in [0, \infty)$ and get

$$f'(t) = t^{q-1} - 1 \begin{cases} < 0, & t \in [0, 1) \\ > 0, & t \in (1, \infty) \end{cases},$$

which ensures that $f(t)$ has minimum 0 at $t = 1$, and hence $f(t) \geq 0$ for all $t \in [0, \infty)$ – in particular $f(ba^{1-p}) \geq 0$ for $a, b \in (0, \infty)$. Consequently,

$$ab \leq p^{-1}a^p + q^{-1}b^q \text{ for all } a, b \in [0, \infty).$$

Taking

$$a = |f_1(x)| \left(\int_{\mathbf{R}} |f_1|^p dm_g \right)^{-\frac{1}{p}} \text{ and } b = |f_2(x)| \left(\int_{\mathbf{R}} |f_2|^q dm_g \right)^{-\frac{1}{q}} \text{ for all } x \in \mathbf{R}$$

in the last inequality, we get

$$\frac{|f_1(x)f_2(x)|}{\left(\int_{\mathbf{R}} |f_1|^p dm_g \right)^{\frac{1}{p}} \left(\int_{\mathbf{R}} |f_2|^q dm_g \right)^{\frac{1}{q}}} \leq \frac{|f_1(x)|^p}{p \int_{\mathbf{R}} |f_1|^p dm_g} + \frac{|f_2(x)|^q}{q \int_{\mathbf{R}} |f_2|^q dm_g},$$

whence deriving the required inequality via an integration with respect to m_g . \square

Remark 4.5.5.

(i) The case $p = q = 2$ of the Hölder inequality is the so-called Cauchy-Schwarz's inequality.

(ii) If $p = 1$, $q = p(p-1)^{-1} = \infty$ and $c \in \mathbf{R}$, then

$$\mathbf{R} = \{x \in \mathbf{R} : |f_2(x)| > c\} \cup \{x \in \mathbf{R} : |f_2(x)| \leq c\},$$

and hence as a limiting case of Theorem 4.5.4, the following Hölder inequality is valid:

$$\int_{\mathbf{R}} |f_1 f_2| dm_g \leq \left(\int_{\mathbf{R}} |f_1| dm_g \right) \inf \{c \geq 0 : m_g(\{x \in \mathbf{R} : |f_2(x)| > c\}) = 0\}.$$

(iii) The Hölder inequality has the following generalization: If $p_1, \dots, p_n \in (1, \infty)$ with $\sum_{j=1}^n p_j^{-1} = 1$, and f_1, \dots, f_n are m_g -measurable, then

$$\int_{\mathbf{R}} \prod_{j=1}^n |f_j| dm_g \leq \prod_{j=1}^n \left(\int_{\mathbf{R}} |f_j|^{p_j} dm_g \right)^{\frac{1}{p_j}}.$$

This can be proved by Hölder's inequality for f_1 and $\prod_{j=2}^n f_j$ and induction on

$$\int_{\mathbf{R}} \prod_{j=2}^n |f_j|^{\frac{p_1}{p_1-1}} dm_g.$$

Now, we can establish the following general form of Minkowski's inequality.

Theorem 4.5.6. Let $g_1, g_2 : \mathbf{R} \rightarrow \mathbf{R}$ be increasing and $1 \leq p < \infty$. If $f : \mathbf{R}^2 \rightarrow \mathbf{R}^e$ is $m_{g_1 \otimes g_2}$ -measurable, then

$$\left(\int_{\mathbf{R}} \left(\int_{\mathbf{R}} |f| dm_{g_2} \right)^p dm_{g_1} \right)^{\frac{1}{p}} \leq \int_{\mathbf{R}} \left(\int_{\mathbf{R}} |f|^p dm_{g_1} \right)^{\frac{1}{p}} dm_{g_2}.$$

Proof. The case $p = 1$ follows immediately from the Fubini-Tonell's theorem. Next, let $p \in (1, \infty)$. Suppose that $F(x_1) = \int_{\mathbf{R}} |f(x_1, x_2)| dm_{g_2}(x_2)$. Then a combination of Fubini-Tonelli's theorem and Hölder's inequality gives

$$\begin{aligned} \int_{\mathbf{R}} F^p dm_{g_1} &= \int_{\mathbf{R}} F^{p-1} F dm_{g_1} \\ &= \int_{\mathbf{R}} (F(x_1))^{p-1} \left(\int_{\mathbf{R}} |f(x_1, x_2)| dm_{g_2}(x_2) \right) dm_{g_1}(x_1) \\ &= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} |f(x_1, x_2)| F^{p-1}(x_1) dm_{g_1}(x_1) \right) dm_{g_2}(x_2) \\ &\leq \int_{\mathbf{R}} \left(\int_{\mathbf{R}} |f(x_1, x_2)|^p dm_{g_1}(x_1) \right)^{\frac{1}{p}} \left(\int_{\mathbf{R}} F^p dm_{g_1} \right)^{p-1} dm_{g_2}(x_2), \end{aligned}$$

whence deriving the required inequality. \square

Remark 4.5.7. In Theorem 4.5.6, if $f(x_1, 1) = f_1(x_1)$ and $f(x_1, 2) = f_2(x_1)$ are m_{g_1} -measurable, and if g_2 is the following step function

$$g_2(x_2) = \begin{cases} 0, & x_2 \in (-\infty, 1) \\ 1, & x_2 \in [1, 2] \\ 2, & x_2 \in (2, \infty) \end{cases},$$

then the well-known Minkowski inequality of the triangle type:

$$\left(\int_{\mathbf{R}} |f_1 + f_2|^p dm_{g_1} \right)^{\frac{1}{p}} \leq \left(\int_{\mathbf{R}} |f_1|^p dm_{g_1} \right)^{\frac{1}{p}} + \left(\int_{\mathbf{R}} |f_2|^p dm_{g_1} \right)^{\frac{1}{p}}$$

holds for $p \in [1, \infty)$ by Remark 4.1.10 (ii) (a)-(b). In case of $p = \infty$, we can readily get

$$\begin{aligned} \inf \{c \geq 0 : m_g(\{x \in \mathbf{R} : |f_1(x) + f_2(x)| > c\}) = 0\} \\ \leq \inf \{c \geq 0 : m_g(\{x \in \mathbf{R} : |f_1(x)| > c\}) = 0\} \\ + \inf \{c \geq 0 : m_g(\{x \in \mathbf{R} : |f_2(x)| > c\}) = 0\}. \end{aligned}$$

Furthermore, it is easy to verify that

$$\int_{\mathbf{R}} |f_1 + f_2|^p dm_{g_1} \leq \int_{\mathbf{R}} |f_1|^p dm_{g_1} + \int_{\mathbf{R}} |f_2|^p dm_{g_1}$$

holds for $p \in (0, 1)$.

Based on Remark 4.5.7 and Definition 4.1.9, we consider the class of Lebesgue-Radon-Stieltjes p -integral functions.

Definition 4.5.8. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing and $p \in (0, \infty]$.

(i) When $p \in (0, \infty)$, the so-called Lebesgue-Radon-Stieltjes space $LRS_g^p(\mathbf{R})$, often denoted $L^p(m_g, \mathbf{R})$, comprises all functions $f : \mathbf{R} \rightarrow \mathbf{R}^e$ satisfying $|f|^p \in LRS_g(\mathbf{R})$ or $S_g(|f|^p) < \infty$. Regarding the limiting case $p = \infty$, $LRS_g^\infty(\mathbf{R})$, often denoted $L^\infty(m_g, \mathbf{R})$, stands for the class of all functions $f : \mathbf{R} \rightarrow \mathbf{R}^e$ with

$$\inf \{c \geq 0 : m_g(\{x \in \mathbf{R} : |f(x)| > c\}) = 0\} < \infty.$$

(ii) If $E \subseteq \mathbf{R}$ is m_g -measurable, then $LRS_g^p(E)$ is defined to be the class of all functions $f : \mathbf{R} \rightarrow \mathbf{R}^e$ with $f1_E \in LRS_g^p(\mathbf{R})$.

Remark 4.5.9. For the purpose of integration it does not make any difference whenever one changes function values on m_g -null sets. Actually, one can integrate functions f that are only defined on an m_g -measurable set E for which $m_g(E^c) = 0$ simply by taking f to be 0 on $E^c = \mathbf{R} \setminus E$. In this manner, extended real-valued functions that are finite m_g -a.e can be treated freely as real-valued functions.

Minkowski's inequality in Remark 4.5.7 is used to produce the following convergence theorem in $LRS_g^p(\mathbf{R})$.

Theorem 4.5.10. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing, $p \in [1, \infty)$, and $C_0(\mathbf{R})$ the class of all continuous functions on \mathbf{R} vanishing outside a bounded and closed subset of \mathbf{R} .*

(i) *If*

$$\begin{cases} f_j, f \in LRS_g^p(\mathbf{R}); \\ \lim_{j \rightarrow \infty} S_g(|f_j - f|^p) = 0; \\ h \in LRS_g^{\frac{p}{p-1}}(\mathbf{R}), \end{cases}$$

then

$$\begin{cases} \lim_{j \rightarrow \infty} S_g(|f_j|^p) = S_g(|f|^p); \\ \lim_{j \rightarrow \infty} S_g(|f_j h - f h|^p) = 0; \\ \lim_{j \rightarrow \infty} \int_{\mathbf{R}} f_j h dm_g = \int_{\mathbf{R}} f h dm_g. \end{cases}$$

(ii) *If*

$$f_j \in LRS_g^p(\mathbf{R}) \text{ and } \lim_{j,k \rightarrow \infty} S_g(|f_j - f_k|^p) = 0,$$

then there exist a function $f \in LRS_g^p(\mathbf{R})$ and a subsequence $\{f_{j_n}\}_{n=1}^\infty$ such that

$$\lim_{j \rightarrow \infty} S_g(|f_j - f|^p) = 0 \text{ and } \lim_{n \rightarrow \infty} f_{j_n}(x) = f(x) \text{ for } m_g - a.e. \ x \in \mathbf{R}.$$

(iii) *If $f \in LRS_g^p(\mathbf{R})$ and $\varepsilon \in (0, 1)$, then there exists a function $h \in C_0(\mathbf{R})$ such that $S_g(|f - h|^p) < \varepsilon^p$.*

Proof. (i) It suffices to handle the case $p \in (1, \infty)$ since $p(p-1)^{-1}$ is treated as ∞ when $p = 1$. Minkowski's inequality in Remark 4.5.7 gives

$$\begin{cases} (S_g(|f_j|^p))^{\frac{1}{p}} \leq (S_g(|f_j - f|^p))^{\frac{1}{p}} + (S_g(|f|^p))^{\frac{1}{p}}; \\ (S_g(|f|^p))^{\frac{1}{p}} \leq (S_g(|f_j - f|^p))^{\frac{1}{p}} + (S_g(|f_j|^p))^{\frac{1}{p}}, \end{cases}$$

whence deriving the first limit result. The second and third limit assertions follow from an application of Hölder's inequality to

$$\begin{aligned} \left| \int_{\mathbf{R}} f_j h dm_g - \int_{\mathbf{R}} f h dm_g \right| &\leq \int_{\mathbf{R}} |f_j - f| |h| dm_g \\ &\leq (S_g(|f_j - f|^p))^{\frac{1}{p}} (S_g(|h|^{\frac{p}{p-1}}))^{\frac{p-1}{p}}. \end{aligned}$$

(ii) Choose $\{j_n\}_{n=1}^\infty$ in \mathbf{N} such that

$$(S_g(|f_j - f_{j_n}|^p))^{\frac{1}{p}} \leq 2^{-n} \text{ for all } j \geq j_n.$$

Then $F_n = \sum_{k=1}^{n-1} |f_{j_{k+1}} - f_{j_k}|$ increases with $n \geq 2$ and hence by Minkowski's inequality in Remark 4.5.7,

$$(S_g(|F_n|^p))^{\frac{1}{p}} \leq \sum_{k=1}^{n-1} (S_g(|f_{j_{k+1}} - f_{j_k}|^p))^{\frac{1}{p}} \leq \sum_{k=1}^{n-1} 2^{-k} < 1.$$

By Fatou's lemma (i.e. Lemma 4.3.3) there is an $F \in LRS_g^p(\mathbf{R})$ such that

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \text{ for } m_g - \text{a.e. } x \in \mathbf{R}.$$

Accordingly, $f_{j_1} + \sum_{k=1}^\infty (f_{j_{k+1}} - f_{j_k})$ is absolutely convergent and its partial sum sequence $\{f_{j_n}\}_{n=1}^\infty$ converges m_g -a.e. on \mathbf{R} to a function $f \in LRS_g^p(\mathbf{R})$ thanks to

$$|f_{j_n}| \leq F_n \leq F \text{ } m_g - \text{a.e. on } \mathbf{R}.$$

A further application of the Minkowski inequality in Remark 4.5.7 gives

$$\begin{aligned} (S_g(|f_j - f|^p))^{\frac{1}{p}} &\leq (S_g(|f_j - f_{j_n}|^p))^{\frac{1}{p}} + (S_g(|f_{j_n} - f|^p))^{\frac{1}{p}} \\ &\leq 2^{-n} + \sum_{k=n}^\infty (S_g(|f_{j_{k+1}} - f_{j_k}|^p))^{\frac{1}{p}} \\ &\leq 2^{2-n}, \end{aligned}$$

which deduces the desired conclusion.

(iii) For $f \in LRS_g^p(\mathbf{R})$ and $j \in \mathbf{N}$ let

$$f_j(x) = \begin{cases} -j, & f(x) < -j \\ f(x), & |f(x)| \leq j \\ j, & f(x) \geq j \end{cases}.$$

Then

$$|f_j(x) - f(x)| \leq |f(x)| \text{ and } \lim_{j \rightarrow \infty} |f_j(x) - f(x)| = 0 \text{ for } m_g - \text{a.e. } x \in \mathbf{R},$$

and hence for any $\varepsilon \in (0, 1)$ there is an $j \in \mathbf{N}$ such that $S_g(|f_j - f|^p) < 2^{-1}\varepsilon$ due to Theorem 4.3.1, and at the same time, there is a step function s on \mathbf{R} such that

$$(S_g(|s - f_j|^p))^{\frac{1}{p}} \leq (2j)^{1-\frac{1}{p}} \left(\int_{\mathbf{R}} |s - f_j| dm_g \right)^{\frac{1}{p}} \leq 2^{-1}\varepsilon.$$

Accordingly, the Minkowski's inequality in Remark 4.5.7 yields

$$(S_g(|f-s|^p))^{\frac{1}{p}} \leq (S_g(|f-f_j|^p))^{\frac{1}{p}} + (S_g(|f_j-s|^p))^{\frac{1}{p}} < \varepsilon.$$

Note that $s = \sum_{k=1}^n c_k 1_{I_k}$ where $\{c_k\}_{k=1}^n$ are constants and $\{I_k\}_{k=1}^n$ are disjoint finite intervals in \mathbf{R} . So, if each 1_{I_k} can be approximated by $C_0(\mathbf{R})$ -functions, then the argument is finished. Without loss of generality (see also the argument for Theorem 2.4.13), we may assume $I_k = [0, 1]$. For any $\varepsilon > 0$ let

$$\phi_\varepsilon(x) = \begin{cases} 0, & x \in (-\infty, -\varepsilon) \\ 1 + \varepsilon^{-1}x, & x \in [-\varepsilon, 0) \\ 1, & x \in [0, 1] \\ \varepsilon^{-1}(1 + \varepsilon - x), & x \in (1, 1 + \varepsilon] \\ 0, & x \in (1 + \varepsilon, \infty) \end{cases}.$$

Then $\phi \in C_0(\mathbf{R})$ and

$$1_{[0,1]} \leq \phi \text{ and } \int_{\mathbf{R}} |\phi_\varepsilon - 1_{[0,1]}|^p dm_g \leq m_g([-\varepsilon, 0)) + m_g((1, 1 + \varepsilon]).$$

Note that the right-hand side of the last inequality tends to zero as ε approaches zero since $\cap_{\varepsilon>0} [-\varepsilon, 0) = \emptyset = \cap_{\varepsilon>0} (1, 1 + \varepsilon]$. So, the desired conclusion follows. \square

The next application of Fubini-Tonelli's theorem is to provide a very useful way evaluating a Lebesgue-Radon-Stieltjes integral via the classical Lebesgue integral or reducing a problem about an integral of a general function to a problem about the integration of characteristic functions.

Theorem 4.5.11. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing and $p \in (0, \infty)$. If $f : \mathbf{R} \rightarrow \mathbf{R}^e$ is m_g -measurable, then*

$$\int_{\mathbf{R}} |f|^p dm_g = p \int_0^\infty m_g(\{x \in \mathbf{R} : |f(x)| > t\}) t^{p-1} dt.$$

Proof. Using Fubini-Tonelli's theorem we have

$$\begin{aligned}
 \int_{\mathbf{R}} |f|^p dm_g &= \int_{\mathbf{R}} \left(\int_0^{|f|} dt^p \right) dm_g \\
 &= p \int_{\mathbf{R}} \left(\int_0^\infty 1_{\{0 < t < |f|\}} t^{p-1} dt \right) dm_g \\
 &= p \int_0^\infty \left(\int_{\mathbf{R}} 1_{\{x \in \mathbf{R} : |f(x)| > t\}} dm_g \right) t^{p-1} dt \\
 &= p \int_0^\infty t^{p-1} m_g(\{x \in \mathbf{R} : |f(x)| > t\}) dt.
 \end{aligned}$$

□

Remark 4.5.12. For a fixed point $c \in \mathbf{R}$, let

$$g(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}.$$

Then Remark 4.1.10 (ii) (b) and Theorem 4.5.11 are employed to deduce the well-known layer cake representation of a function $f : \mathbf{R} \rightarrow [0, \infty]$:

$$f(c) = \int_0^\infty 1_{\{x \in \mathbf{R} : f(x) > t\}}(c) dt.$$

Even more interesting and important is that Theorem 4.5.11 suggest a useful approach to define an integral on an abstract set.

Definition 4.5.13. For a given set X let $\mathcal{P}(X)$ be the collection of all subsets of X and $\mu : \mathcal{P}(X) \rightarrow [0, \infty]$ an increasing set function; that is, μ satisfies both $\mu(\emptyset) = 0$ and $\mu(E_1) \leq \mu(E_2)$ for $E_1 \subseteq E_2 \subseteq X$. Suppose $f : X \rightarrow [0, \infty]$. Then $\mu(\{x \in X : f(x) > t\})$ is an increasing function of $t \in (0, \infty)$, and hence

$$\int_X f d\mu = \int_0^\infty \mu(\{x \in X : f(x) > t\}) dt$$

can be used to define the integral of f on X . In general, for $f : X \rightarrow \mathbf{R}^e$ we write

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu.$$

Theorem 4.5.14. The integral introduced in Definition 4.5.13 enjoys the following four properties.

- (i) $f_1, f_2 : X \rightarrow [0, \infty]$ and $f_1 \leq f_2$ implies $\int_X f_1 d\mu \leq \int_X f_2 d\mu$.
(ii) $\int_X \min\{f, c\} d\mu + \int_X (f - c)^+ d\mu = \int_X f d\mu$ for any $c \in [0, \infty)$.
(iii) $\int_X f d\mu = \lim_{\varepsilon \rightarrow 0^+} \int_X \min\{(f - \varepsilon)^+, \varepsilon^{-1}\} d\mu$.
(iv) $\int_X c f d\mu = c \int_X f d\mu$ for $c \in [0, \infty)$.

Conversely, if a real-valued function L defined on the family of functions from $\mathcal{P}(X)$ to $[0, \infty]$ obeys (i)-(ii)-(iii)-(iv) above, and $\mu(E) = L(1_E)$ is defined for any $E \in \mathcal{P}(X)$, then

$$L(f) = \int_X f d\mu \text{ for all } f : X \rightarrow [0, \infty].$$

Proof. (i) This follows from the definition.

(ii) This follows from the following two formulas:

$$\begin{cases} \int_X \min\{f, c\} d\mu = \int_0^c \mu(\{x \in X : f(x) > t\}) dt; \\ \int_X (f - c)^+ d\mu = \int_c^\infty \mu(\{x \in X : f(x) > t\}) dt. \end{cases}$$

(iii) This follows from the monotone property of the function

$$t \mapsto \mu(\{x \in X : f(x) > t\})$$

and the following calculation

$$\int_X \min\{(f - \varepsilon)^+, \varepsilon^{-1}\} d\mu = \int_\varepsilon^{\varepsilon + \varepsilon^{-1}} \mu(\{x \in X : f(x) > t\}) dt.$$

(iv) This follows from the substitution $t = cs$.

For the reversed result, suppose that L satisfies (i)-(ii)-(iii)-(iv) and set $\mu(E) = L(1_E)$ for $E \in \mathcal{P}(X)$. Fix $f : X \rightarrow [0, \infty]$. Then

$$f \geq \varepsilon 1_{\{x \in X : f(x) > \varepsilon\}} \text{ for any } \varepsilon > 0,$$

and hence we may assume that $\mu(\{x \in X : f(x) > \varepsilon\})$ is finite – otherwise it follows from (i) that

$$L(f) \geq \varepsilon L(1_{\{x \in X : f(x) > \varepsilon\}}) = \infty = \int_X f d\mu$$

and so that the above inequality becomes an equality.

Given $\varepsilon > 0$ and $n \in \mathbf{N}$, set

$$\begin{cases} c = \varepsilon + \varepsilon^{-1}; \\ T_\varepsilon(f) = \min\{(f - \varepsilon)^+, \varepsilon^{-1}\}; \\ X_{n,j} = \{x \in X : T_\varepsilon(f) > jcn^{-1}\} \text{ for } j \in \{0, \dots, n-1\}. \end{cases}$$

Using (ii) we obtain

$$\begin{aligned} L(T_\varepsilon(f)) - \int_X T_\varepsilon(f) d\mu \\ = \sum_{j=0}^{n-1} (L(\min\{T_\varepsilon(f) - jcn^{-1}, cn^{-1}1_{X_{n,j}}\}) - \int_X \min\{T_\varepsilon(f) - jcn^{-1}, cn^{-1}1_{X_{n,j}}\} d\mu). \end{aligned}$$

Meanwhile, using (i) and (iv) we find the following two estimates:

$$\begin{cases} cn^{-1}\mu(X_{n,j+1}) \leq L(\min\{T_\varepsilon(f) - jcn^{-1}, cn^{-1}1_{X_{n,j}}\}) \leq cn^{-1}\mu(X_{n,j}); \\ cn^{-1}\mu(X_{n,j+1}) \leq \int_X (\min\{T_\varepsilon(f) - jcn^{-1}, cn^{-1}1_{X_{n,j}}\}) d\mu \leq cn^{-1}\mu(X_{n,j}). \end{cases}$$

Therefore

$$\left| L(T_\varepsilon(f)) - \int_X T_\varepsilon(f) d\mu \right| \leq cn^{-1}\mu(X_{n,0}) = cn^{-1}\mu(\{x \in X : f(x) > \varepsilon\}).$$

Since $\varepsilon > 0$ and $n \in \mathbf{N}$ are arbitrary, this last estimate, along with (iii), must ensure $L(f) = \int_X f d\mu$. \square

As an aside of Theorem 4.5.14, we can establish the following properties of the above-defined integrals.

Corollary 4.5.15. *For a set X let $\mu : \mathcal{P}(X) \rightarrow [0, \infty]$ be an increasing set function, $f : X \rightarrow [0, \infty]$ and $E \subseteq X$.*

(i) *If $c \in [0, \infty)$, then*

$$\int_X (f + c)1_E d\mu = \int_X f1_E d\mu + c\mu(E).$$

(ii) *If μ is sub-additive; that is,*

$$\mu(E_1 \cup E_2) \leq \mu(E_1) + \mu(E_2) \text{ for all } E_1, E_2 \subseteq X,$$

then

$$\int_X f1_S d\mu \leq \int_X f1_{S \cap E} d\mu + \int_X f1_{S \setminus E} d\mu \text{ for all } S \subseteq X$$

with equality if E is μ -measurable; that is,

$$\mu(A) = \mu(A \cap E) + \mu(A \setminus E) \text{ for all } A \subseteq X.$$

Proof. (i) This follows from Theorem 4.5.14 (ii) applied to

$$\min\{(f+c)1_E, c\} = c1_E \quad \text{and} \quad ((f+c)1_E - c)^+ = f1_E.$$

(ii) Using both the sub-additivity of μ and the equality $S = (S \cap E) \cup (S \setminus E)$ for any $S \subseteq X$, we have

$$\begin{aligned} \int_X f1_S d\mu &= \int_0^\infty \mu(\{x \in X : 1_S(x)f(x) > t\}) dt \\ &\leq \int_0^\infty \mu(\{x \in X : 1_S(x)f(x) > t\}) dt \\ &\quad + \int_0^\infty \mu(\{x \in X : 1_{S \setminus E}(x)f(x) > t\}) dt \\ &= \int_X f1_{S \cap E} d\mu + \int_X f1_{S \setminus E} d\mu, \end{aligned}$$

where equality occurs when E is μ -measurable since $1_S = 1_{S \cap E} + 1_{S \setminus E}$. \square

Remark 4.5.16. We can develop a measure theoretic based theory of integration via establishing countable subadditivity/additivity, dominated convergence theorem, monotone convergence theorem and Fatou's lemma, Fubini-Tonelli's theorem and so on.

Problems

4.1. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be determined by

$$g(x) = \begin{cases} x, & x < 0 \\ 2, & x = 0 \\ 3 - e^{-x}, & x > 0 \end{cases}.$$

Find $m_g(I)$ when $I = (0, 1); [0, 1]; (-1, 1); [0, 0]$.

4.2. Suppose $g : \mathbf{R} \rightarrow \mathbf{R}$ is an increasing function.

(i) Prove that if S_1 and S_2 are disjoint simple subsets of \mathbf{R} then $m_g(S_1 \cup S_2) = m_g(S_1) + m_g(S_2)$. Construct examples to show that if S_1 and S_2 are not disjoint, then $m_g(S_1 \cup S_2)$ may or may not equal $m_g(S_1) + m_g(S_2)$.

(ii) Prove that if S_1 and S_2 are simple subsets of \mathbf{R} such that $S_1 \subseteq S_2$ then $m_g(S_2 \setminus S_1) = m_g(S_2) - m_g(S_1)$. Construct examples to show that if S_1 and S_2 are simple with $S_1 \not\subseteq S_2$ then $m_g(S_2 \setminus S_1)$ may or may not equal $m_g(S_2) - m_g(S_1)$.

4.3. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be given by

$$g(x) = \begin{cases} 2^{-1}x, & x < 0 \\ 1, & x \geq 0 \end{cases}.$$

Decide whether or not the following five functions are step functions on \mathbf{R} . If the answer is yes, then find the corresponding $A_g(\cdot)$:

(i)

$$s_1(x) = \begin{cases} 0, & x \in \mathbf{R} \setminus (-\infty, -1) \\ 1, & x \in [-1, 0) \end{cases};$$

(ii)

$$s_2(x) = \begin{cases} -2, & x \in [-1, 0) \\ 1, & x \in [0, 1] \\ 0, & x \in [-1, 1] \end{cases};$$

(iii)

$$s_3(x) = \begin{cases} 2, & x \in [-1, 3] \\ 1, & x \in (3, \infty) \\ 0, & x \in \mathbf{R} \setminus (-\infty, -1) \end{cases};$$

(iv)

$$s_4(x) = \begin{cases} 3, & x \in (-2, -1) \\ -3, & x \in [-1, 1) \\ 0, & x \in \mathbf{R} \setminus (-2, 1) \end{cases};$$

(v)

$$s_5(x) = \begin{cases} -1, & x \in (-\infty, 0) \\ 1, & x \in [0, \infty) \end{cases}.$$

4.4. Give a sequence of step functions $\{s_k\}_{k=1}^{\infty}$ on \mathbf{R} such that $f = \sum_{k=1}^{\infty} s_k$ is not a step function and $S_{id}(f) = 0$.

4.5. Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be the following function

$$f(x) = \begin{cases} n^{-1}, & x \in [n-1, n-2^{-1}), n \in \mathbf{N} \\ -n^{-1}, & x \in [n-2^{-1}, n), n \in \mathbf{N} \\ 0, & x \in (-\infty, 0) \end{cases}.$$

(i) Evaluate $\int_0^b f(x) dx$ for any $b > 0$.

(ii) Prove that $\int_0^{\infty} f(x) dx$ exists.

(iii) Prove that $\lim_{b \rightarrow \infty} \int_0^b |f(x)| dx$ and $\int_{\mathbf{R}} 1_{[0, \infty)} f dm_{id}$ do not exist.

4.6. Prove the following mean value inequality for the Lebesgue-Radon-Stieltjes integrals. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing and $I \subseteq \mathbf{R}$ be a finite interval. If $1_I f \in LRS_g(\mathbf{R})$ and $c_1 \leq f \leq c_2$ on I for some finite constants c_1, c_2 , then $c_1 m_g(I) \leq \int_{\mathbf{R}} f 1_I dm_g \leq c_2 m_g(I)$.

4.7.

(i) Given an increasing function $g : \mathbf{R} \rightarrow \mathbf{R}$. Show that if $f_1, f_2, \dots, f_n \in LRS_g(\mathbf{R})$ then

$$\left| \int_{\mathbf{R}} \left(\sum_{k=1}^n f_k \right) dm_g \right| \leq \sum_{k=1}^n \int_{\mathbf{R}} |f_k| dm_g.$$

(ii) Let $g = \sum_{k=1}^n c_k g_k$, where each $g_k : \mathbf{R} \rightarrow \mathbf{R}$ is increasing and each $c_k \geq 0$. Prove that

$$f \in \cap_{k=1}^n LRS_{g_k}(\mathbf{R}) \Rightarrow f \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} f dm_g = \sum_{k=1}^n c_k \int_{\mathbf{R}} f dm_{g_k}.$$

4.8.

(i) Show that the union of two m_g -null sets is an m_g -null set.

(ii) Give an example of a finite subset $E \subseteq \mathbf{R}$ and an increasing function $g : \mathbf{R} \rightarrow \mathbf{R}$ such that E is not an m_g -null set.

(iii) Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be an increasing function. Verify that if $f_1, f_2 \in LRS_g(\mathbf{R})$ and $f_1 \leq f_2$ m_g -a.e. on \mathbf{R} , then $\int_{\mathbf{R}} f_1 dm_g \leq \int_{\mathbf{R}} f_2 dm_g$.

4.9. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be an increasing function. Suppose that $\{f_k\}_{k=1}^{\infty}$ is a sequence in $LRS_g(\mathbf{R})$ and $\sum_{k=1}^{\infty} f_k = f$ exists m_g -a.e. on \mathbf{R} . Prove the following results.

(i) If there is a function $F \in LRS_g(\mathbf{R})$ such that $|\sum_{k=1}^n f_k| \leq F$ on \mathbf{R} , then

$$f \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} f dm_g = \sum_{k=1}^{\infty} \int_{\mathbf{R}} f_k dm_g.$$

(ii) If $f_k \geq 0$ or $f_k \leq 0$ on \mathbf{R} and $\sum_{k=1}^{\infty} \int_{\mathbf{R}} f_k dm_g$ converges, then

$$f \in LRS_g(\mathbf{R}) \text{ and } \int_{\mathbf{R}} f dm_g = \sum_{k=1}^{\infty} \int_{\mathbf{R}} f_k dm_g.$$

4.10. For $n \in \mathbf{N}$ let

$$f_n(x) = \begin{cases} 1, & x \in [n, n+1) \\ 0, & x \in \mathbf{R} \setminus [n, n+1) \end{cases}.$$

(i) Find $f = \lim_{n \rightarrow \infty} f_n$.

(ii) Prove

$$\lim_{n \rightarrow \infty} \int_{\mathbf{R}} f_n dm_{id} \neq \int_{\mathbf{R}} f dm_{id}.$$

(iii) Point out which hypothesis of the monotone convergence theorem is not valid in this case, and write down the reasoning why it is not valid.

4.11. If $\alpha > 0$, prove the following two formulas:

(i)

$$\lim_{n \rightarrow \infty} \int_{\mathbf{R}} 1_{[0, n]}(x) (1 - n^{-1}x)^n |x|^{\alpha-1} dm_{id}(x) = \int_{\mathbf{R}} 1_{(0, \infty)}(x) e^{-x} |x|^{\alpha-1} dm_{id}(x);$$

(ii)

$$\sum_{n=1}^{\infty} \int_{\mathbf{R}} 1_{(0, \infty)}(x) \left(\frac{|x|^\alpha}{e^{nx}} \right) dm_{id}(x) = \int_{\mathbf{R}} 1_{(0, \infty)}(x) (e^x - 1) |x|^\alpha dm_{id}(x) = \Gamma(1 + \alpha) \sum_{n=1}^{\infty} n^{-1-\alpha}.$$

4.12. Prove the following formulas:

(i)

$$\int_{\mathbf{R}} 1_{(0, 1]}(x) |x|^{-p} dm_{id}(x) = \begin{cases} (1-p)^{-1}, & p \in (0, 1) \\ \infty, & p \in [1, \infty) \end{cases};$$

(ii)

$$\lim_{n \rightarrow \infty} \int_{\mathbf{R}} 1_{(0, \infty)}(x) |x|^{-\frac{1}{n}} (1 + n^{-1}|x|)^{-n} dm_{id}(x) = 1.$$

4.13. For each $n \in \mathbf{N}$ and all $x \in \mathbf{R}$, let $f_n(x) = e^{-nx} - 2e^{-2nx}$.

(i) Prove that $f(x) = \sum_{n=1}^{\infty} f_n(x)$ is convergent for all $x > 0$, and calculate $f(x)$.

(ii) Prove that each $1_{(0, \infty)} f_n$ and $1_{(0, \infty)} f$, is Lebesgue integrable on \mathbf{R} .

(iii) Compare $\int_{\mathbf{R}} 1_{(0, \infty)}(x) f(x) dm_{id}(x)$ and $\sum_{n=1}^{\infty} \int_{\mathbf{R}} 1_{(0, \infty)}(x) f_n(x) dm_{id}(x)$.

4.14.

(i) For each $n \in \mathbf{N}$ let

$$f_n = \begin{cases} 1_{[0, \frac{1}{4}]} & , \ n \text{ is even} \\ 1_{[\frac{1}{4}, 1]} & , \ n \text{ is odd} \end{cases}.$$

Calculate the following four numbers:

$$\begin{cases} N_1 = \lim_{k \rightarrow \infty} \inf_{n \geq k} \int_0^\infty f_n(x) dx; \\ N_2 = \int_0^\infty \lim_{k \rightarrow \infty} \inf_{n \geq k} f_n(x) dx; \\ N_3 = \lim_{k \rightarrow \infty} \sup_{n \geq k} \int_0^\infty f_n(x) dx; \\ N_4 = \int_0^\infty \lim_{k \rightarrow \infty} \sup_{n \geq k} f_n(x) dx. \end{cases}$$

(ii) If $\{f_n\}_{n=1}^\infty$ is a sequence of positive Lebesgue measurable function, what can be said of the four numbers above, and more particularly about N_3 and N_4 ?

4.15. Let $g(x) = x$ on \mathbf{R} . If $f \in LRS_g(0, \infty)$ is uniformly continuous on $(0, \infty)$, prove $\lim_{x \rightarrow \infty} f(x) = 0$.

4.16. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing and

$$\begin{cases} f : [a, b] \times [c, d] \rightarrow \mathbf{R}; \\ y \in [c, d]; \\ f(\cdot, y) \in LRS_g[a, b]; \\ F(y) = \int_{[a, b]} f(x, y) dm_g(x). \end{cases}$$

(i) If there is an $h \in LRS_g([a, b])$ such that $|f(x, y)| \leq h(x)$ for all $(x, y) \in [a, b] \times [c, d]$, and if $\lim_{y \rightarrow y_0} f(x, y) = f(x, y_0)$ for all $(x, y_0) \in [a, b] \times [c, d]$, then prove that $\lim_{y \rightarrow y_0} F(y) = F(y_0)$ - in particular - if $f(x, \cdot)$ is continuous for every $x \in [a, b]$, then F is continuous.

(ii) If $\partial_y f(x, y)$ exists and there is an $h \in LRS_g([a, b])$ such that

$$|\partial_y f(x, y)| \leq h(x) \text{ for all } (x, y) \in [a, b] \times [c, d],$$

then prove that F is differentiable and $F'(y) = \int_{[a, b]} \partial_y f(\cdot, y) dm_g(\cdot)$ for any $y \in [c, d]$.

4.17. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing and \mathcal{M}_g be the σ -ring of all m_g -measurable sets. Prove that if $\{E_j\}_{j=1}^\infty$ is a sequence of sets in \mathcal{M}_g then $m_g(\cup_{j=1}^\infty E_j) \leq \sum_{j=1}^\infty m_g(E_j)$.

4.18. Given an increasing function $g : \mathbf{R} \rightarrow \mathbf{R}$. For an m_g -measurable set $E \subseteq \mathbf{R}$ let $L_g(E) = \int_E dm_g < \infty$. If $1_E f : \mathbf{R} \rightarrow \mathbf{R}^e$ is m_g -measurable and $E_j = \{x \in E : j-1 \leq f(x) < j\}$ for each $j \in \mathbf{Z}$, prove that

$$\int_E |f| dm_g < \infty \Leftrightarrow \sum_{j=-\infty}^{\infty} |j| m_g(E_j) < \infty.$$

4.19. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. Prove that $f : \mathbf{R} \rightarrow \mathbf{R}^e$ is m_g -measurable if and only if $\{x \in \mathbf{R} : f(x) \geq c\}$ is m_g -measurable for any $c \in \mathbf{R}$.

4.20. Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing. A sequence $\{f_n\}_{n=1}^\infty$ of m_g -measurable functions on \mathbf{R} is said to be:

(i) convergent to f in the Lebesgue-Radon-Stieltjes g -measure provided that

$$\lim_{n \rightarrow \infty} m_g(\{x \in \mathbf{R} : |f_n(x) - f(x)| \geq \varepsilon\}) = 0 \text{ for any } \varepsilon > 0;$$

(ii) a Cauchy sequence in the Lebesgue-Radon-Stieltjes g -measure provided that

$$\lim_{k, n \rightarrow \infty} m_g(\{x \in \mathbf{R} : |f_k(x) - f_n(x)| \geq \varepsilon\}) = 0 \text{ for any } \varepsilon > 0.$$

Prove the following results.

(iii) If $\lim_{n \rightarrow \infty} \int_{\mathbf{R}} |f_n - f| dm_g = 0$, then $f_n \rightarrow f$ in the Lebesgue-Radon-Stieltjes g -measure m_g .

(iv) If $\{f_n\}_{n=1}^\infty$ is a Cauchy sequence in the Lebesgue-Radon-Stieltjes g -measure, then there exists an m_g -a.e. unique function f such that $\{f_n\}_{n=1}^\infty$ is convergent to f in m_g , and hence there exists a subsequence $\{f_{n_k}\}_{k=1}^\infty$ of $\{f_n\}_{n=1}^\infty$ which is convergent to f m_g -a.e. on \mathbf{R} .

(v) If $\{f_n\}_{n=1}^\infty$ is a sequence of nonnegative m_g -measurable functions which converges to f m_g -a.e. on \mathbf{R} , then $\lim_{n \rightarrow \infty} \int_{\mathbf{R}} f_n dm_g = 0 \Rightarrow f = 0$ m_g -a.e. on \mathbf{R} .

(vi) Give an example of g to show that $f_n \rightarrow f$ m_g -a.e. on \mathbf{R} does not ensure $f_n \rightarrow f$ in the g -measure m_g .

4.21. For $p \in (0, 1)$ let

$$f_1(x) = |x|^{-p} 1_{\{x \in \mathbf{R} : |x| < 1\}} \quad \text{and} \quad f_2(x) = (-\ln |x|) 1_{\{x \in \mathbf{R} : |x| < 1\}}.$$

Calculate $\int_{\mathbf{R}} f_j dm_{id}$ for $j = 1, 2$ in two ways:

- (i) Use the standard calculus method;
- (ii) Compute $m_{id}(\{x \in \mathbf{R} : f_j(x) > t\})$.

Chapter 5

Absolute Continuities in Lebesgue Integrals

In Differential-Integral Analysis, the notion of absolute continuity is a smoothness nature of functions which is stronger than the uniform continuity which leads to the Lebesgue differentiation theorem. The absolute continuity allows us to generalize the link between the two fundamental operations within calculus - differentials and integrals. The link is usually characterized by the fundamental theorem of calculus under the Riemann integrals, but nevertheless with the absolute continuity the link may be formulated by means of the Lebesgue integrals. This issue will be addressed in the coming-up-next five sections.

5.1 Lebesgue's Outer Measure and Vitali's Covering

We start with characterizing the structure of an m_{id} -measurable subset of \mathbf{R} .

Theorem 5.1.1. *Let $E \subseteq [a, b]$ be m_{id} -measurable and $\varepsilon > 0$. Then:*

- (i) *there is a union F of finitely many intervals such that $m_{id}((E \setminus F) \cup (F \setminus E)) < \varepsilon$;*
- (ii) *there is an open set $O \supseteq E$ such that $0 \leq m_{id}(O) - m_{id}(E) < \varepsilon$;*
- (iii) *there are an $N \in \mathbf{N}$ and a decreasing sequence of open sets $\{O_j\}_{j=1}^{\infty}$ such that $O_j \supseteq E$ and $0 \leq m_{id}(O_j) - m_{id}(E) < \varepsilon$ when $j \geq N$.*

Proof. (i) Since 1_E is m_{id} -measurable, there is a sequence of step functions $\{\psi_j\}_{j=1}^\infty$ such that $1_E = \lim_{j \rightarrow \infty} \psi_j$ m_{id} -a.e. on E . Upon choosing

$$\phi_j(x) = \begin{cases} 1, & \psi_j(x) \geq 2^{-1} \\ 0, & \psi_j(x) < 2^{-1} \end{cases},$$

we read off that $\lim_{j \rightarrow \infty} \phi_j = 1_E$ m_{id} -a.e. on E . Note that ϕ_j is an indicator 1_{F_j} of a set F_j which is a union of finitely many intervals. So, we have

$$\max\{\phi_j - 1_E, 0\} = 1_{F_j \setminus E}; \quad \max\{1_E - \phi_j, 0\} = 1_{E \setminus F_j}; \quad |\phi_j - 1_E| = 1_{F_j \setminus E} + 1_{E \setminus F_j}.$$

Since

$$\lim_{j \rightarrow \infty} (m_{id}(E \setminus F_j) + m_{id}(F_j \setminus E)) = \lim_{j \rightarrow \infty} \int_{\mathbf{R}} |\phi_j - 1_E| dm_{id} = 0,$$

there is an $n_0 \in \mathbf{N}$ such that

$$m_{id}(E \setminus F_{n_0}) + m_{id}(F_{n_0} \setminus E) < \varepsilon,$$

as desired.

(ii) In accordance with (i), for each $j \in \mathbf{N}$ we can get an open set F_j such that

$$m_{id}((E \setminus F_j) \cup (F_j \setminus E)) < 2^{-j-1} \varepsilon.$$

Upon choosing

$$F_0 = \bigcup_{j=1}^\infty F_j; \quad F_* = \bigcap_{j=1}^\infty (E \setminus F_j); \quad F_\star = \bigcup_{j=1}^\infty (F_j \setminus E),$$

we obtain

$$m_{id}(F_*) = \lim_{j \rightarrow \infty} m_{id}(E \setminus F_j) = 0 \quad \text{and} \quad m_{id}(F_\star) \leq \sum_{j=1}^\infty m_{id}(F_j \setminus E) < 2^{-1} \varepsilon.$$

Note that

$$m_{id}(F_*) = 0 \quad \text{and} \quad F_0 = (E \cap F_0) \cup (F_0 \setminus E) \subseteq E \cup F_\star.$$

So, not only there is an open set $O_0 \supseteq F_\cap$ enjoying $m_{id}(O_0) < 2^{-1} \varepsilon$, but also, $O = F_0 \cup O_0$ is an open set containing E and satisfies

$$m_{id}(O) \leq m_{id}(F_0) + m_{id}(O_0) \leq m_{id}(E) + m_{id}(F_\cup) + m_{id}(O_0) < m_{id}(E) + \varepsilon,$$

as desired.

(iii) According to (ii), for each $j \in \mathbf{N}$ there are an $N \in \mathbf{N}$ and an open set $F_j \supseteq E$ such that

$$j \geq N \Rightarrow 0 \leq m_{id}(F_j) - m_{id}(E) < \varepsilon.$$

If $O_j = \cap_{k=1}^j F_k$, then O_j is open, $O_j \supseteq O_{j+1}$ and

$$j \geq N \Rightarrow 0 \leq m_{id}(F_j) - m_{id}(E) \leq m_{id}(O_j) - m_{id}(E) < \varepsilon,$$

as desired. □

Next, we bring the concepts of Lebesgue's outer measure and Vitali's covering (due to Giuseppe Vitali) into play.

Definition 5.1.2. Let $E \subseteq \mathbf{R}$.

(i) The Lebesgue outer measure $m_{id,*}(E)$ of E is defined as the infimum of $\sum_j m_{id}(I_j)$ for all sequences of open intervals $I_j \subseteq \mathbf{R}$ enjoying $E \subseteq \cup_j I_j$.

(ii) A Vitali's covering \mathcal{V} of E is a family of closed intervals such that for any pair $(\varepsilon, x) \in (0, \infty) \times E$ there is an $I \in \mathcal{V}$ with $0 < m_{id}(I) < \varepsilon$ and $x \in I$.

Theorem 5.1.3. For $j \in \mathbf{N}$ let $E, E_j \subseteq \mathbf{R}$ obey $m_{id,*}(E), m_{id,*}(E_j) < \infty$ and E have a Vitali's covering \mathcal{V} and its $\mathbf{R} \ni t$ -translation $E + t = \{x + t : x \in E\}$. Then

(i)

$$m_{id,*}(E) = \inf\{m_{id}(O) : \text{open } O \supseteq E\} = m_{id,*}(E + t).$$

(ii)

$$\begin{cases} \text{(ii-1)} & E_1 \subseteq E_2 \Rightarrow m_{id,*}(E_1) \leq m_{id,*}(E_2); \\ \text{(ii-2)} & E_j \subseteq E_{j+1} \Rightarrow m_{id,*}(\cup_{j=1}^{\infty} E_j) = \lim_{j \rightarrow \infty} m_{id,*}(E_j); \\ \text{(ii-3)} & m_{id,*}(\cup_{j=1}^{\infty} E_j) \leq \sum_{j=1}^{\infty} m_{id,*}(E_j). \end{cases}$$

(iii) $m_{id,*}(E_0) = m_{id}(E_0)$ for any m_{id} -measurable set $E_0 \subseteq \mathbf{R}$ - in particular - $m_{id,*}(I) = m_{id}(I)$ for any interval $I \subseteq \mathbf{R}$.

(iv) for an arbitrarily $\varepsilon > 0$ there exist mutually disjoint intervals $I_1, \dots, I_n \in \mathcal{V}$ such that $m_{id,*}(E \setminus \cup_{j=1}^n I_j) < \varepsilon$.

Proof. (i) On the one hand, the translation-invariance $m_{id,*}(E) = m_{id,*}(E+t)$ follows from Definition 5.1.2(i) and the translation-invariance $m_{id}(I+t) = m_{id}(I)$ for any interval $I \subseteq \mathbf{R}$.

On the other hand, if $O \subseteq \mathbf{R}$ is an open set containing E , then there exists a sequence $\{I_j\}_{j=1}^\infty$ of mutually disjoint open intervals such that $O = \bigcup_{j=1}^\infty I_j$, and hence

$$\begin{cases} 1_O = \sum_{j=1}^\infty 1_{I_j}; \\ m_{id}(O) = \sum_{j=1}^\infty m_{id}(I_j); \\ m_{id,*}(E) \leq m_{id}(O). \end{cases}$$

Furthermore, for any $\varepsilon > 0$ if $I_j = (a_j, b_j)$, then its ε -enlargement

$$I_{j,\varepsilon} = (a_j - \varepsilon 2^{-j-1}, b_j + \varepsilon 2^{-j-1})$$

produces an open set

$$O_\varepsilon = \bigcup_{j=1}^\infty I_{j,\varepsilon} \supseteq \bigcup_{j=1}^\infty I_j = O \supseteq E$$

such that

$$m_{id}(O_\varepsilon) \leq \sum_{j=1}^\infty m_{id}(I_{j,\varepsilon}) \leq \sum_{j=1}^\infty m_{id}(I_j) + \varepsilon$$

and hence Definition 5.1.2(i) ensures

$$m_{id}(O) \leq m_{id}(O_\varepsilon) \leq m_{id,*}(E) + \varepsilon.$$

Accordingly,

$$m_{id,*}(E) \leq \inf\{m_{id}(O) : \text{open } O \supseteq E\} \leq m_{id,*}(E) + \varepsilon$$

which yields the desired equality.

(ii) Regarding (ii-1), if $E_1 \subseteq E_2$, then any interval covering of E_2 is an interval covering of E_1 , and hence Definition 5.1.2(i) derives $m_{id,*}(E_1) \leq m_{id,*}(E_2)$.

Regarding (ii-2), we use (ii-1) to get that if $E_j \subseteq E_{j+1}$ then $m_{id,*}(E_j) \leq m_{id,*}(E_{j+1})$ and hence

$$\lim_{j \rightarrow \infty} m_{id,*}(E_j) \leq m_{id,*}(\bigcup_{j=1}^\infty E_j).$$

So, if $\lim_{j \rightarrow \infty} m_{id,*}(E_j) = \infty$, then (ii-2) is trivially true. This leads us to considering the case $\lim_{j \rightarrow \infty} m_{id,*}(E_j) < \infty$. Now, for any $\varepsilon > 0$ and $j \in \mathbf{N}$, according to (i) there is a sequence of mutually disjoint open intervals $\{I_{j,k}\}_{k=1}^\infty$ such that

$$E_j \subseteq \bigcup_{k=1}^\infty I_{j,k} = I_j \text{ and } m_{id}(I_j) = \sum_{k=1}^\infty m_{id}(I_{j,k}) < m_{id,*}(E_j) + 2^{-j}\varepsilon.$$

Note that I_j is open and

$$m_{id}(I_1) < m_{id,*}(E_1) + (1 - 2^{-1})\epsilon.$$

So, if $J_j = \bigcap_{k=j}^{\infty} I_k$ (which is m_{id} -measurable, increases and contains E_j), then by induction we may assume

$$m_{id}(\bigcup_{k=1}^j J_k) < m_{id,*}(E_j) + (1 - 2^{-j})\epsilon$$

to achieve (via Theorem 4.4.6(ii))

$$\begin{aligned} m_{id}(\bigcup_{k=1}^{j+1} J_k) &= m_{id}(\bigcup_{k=1}^j J_k) + m_{id}(J_{j+1}) - m_{id}((\bigcup_{k=1}^j J_k) \cap J_{j+1}) \\ &< (m_{id,*}(E_j) + (1 - 2^{-j})\epsilon) + (m_{id,*}(E_{j+1}) + 2^{-j-1}\epsilon) - m_{id,*}(E_j) \\ &= m_{id,*}(E_{j+1}) + (1 - 2^{-j-1})\epsilon. \end{aligned}$$

Consequently, we obtain

$$m_{id}(\bigcup_{k=1}^l J_k) < m_{id,*}(E_l) + (1 - 2^{-l})\epsilon \text{ for all } l \in \mathbf{N},$$

thereby finding via sending $\epsilon \rightarrow 0$,

$$m_{id}(\bigcup_{k=1}^l J_k) \leq m_{id,*}(E_l) \text{ for all } l \in \mathbf{N}.$$

This in turn implies

$$\begin{aligned} m_{id,*}(\bigcup_{j=1}^{\infty} E_j) &\leq m_{id}(\bigcup_{j=1}^{\infty} J_j) \\ &= m_{id}(\bigcup_{k=1}^{\infty} \bigcup_{j=1}^k J_j) \\ &= \lim_{k \rightarrow \infty} m_{id}(\bigcup_{j=1}^k J_j) \\ &\leq \lim_{k \rightarrow \infty} m_{id,*}(E_k). \end{aligned}$$

thereby reaching the desired equality.

Regarding (ii-3), for any $\epsilon > 0$ and each E_j , according to (i) there is a sequence of mutually disjoint intervals $\{I_{j,k}\}_{k=1}^{\infty}$ such that

$$E_j \subseteq \bigcup_{k=1}^{\infty} I_{j,k} \text{ and } \sum_{k=1}^{\infty} m_{id}(I_{j,k}) = m_{id}(\bigcup_{k=1}^{\infty} I_{j,k}) < m_{id,*}(E_j) + \epsilon 2^{-j}.$$

Note that $\bigcup_{j,k=1}^{\infty} I_{j,k} \supseteq \bigcup_{j=1}^{\infty} E_j$. So we have

$$m_{id,*}(\bigcup_{j=1}^{\infty} E_j) \leq \sum_{j,k=1}^{\infty} m_{id}(I_{j,k}) \leq \sum_{j=1}^{\infty} m_{id,*}(E_j) + \epsilon,$$

as desired.

(iii) Clearly, if $E_0 \subseteq \mathbf{R}$ is m_{id} -measurable, then (i) and Theorem 5.1.1(ii) derive

$$m_{id,*}(E_0) = \inf\{m_{id}(O) : \text{open } O \supseteq E_0\} = m_{id}(E_0).$$

However, for the special case that E_0 is an interval $I \subseteq \mathbf{R}$, we offer a direct argument. Note that $m_{id,*}(I) \leq m_{id}(I)$ follows from Definition 5.1.2(i). So, it is required to verify the reversed inequality. To do so, we consider three situations.

The first is that I is a compact (bounded and closed) interval $[a, b]$. If $\{I_j\}_{j=1}^{\infty}$ is a sequence of open intervals covering I , then there is a finite number of open intervals - for example - $\{I_{j_k}\}_{k=1}^n$ such that $I \subseteq \bigcup_{k=1}^n I_{j_k}$, and hence there is an open interval, say, $I_{j_1} = (a_1, b_1)$ such that $a_1 < a < b_1$. Of course, if $b < b_1$, then $I \subseteq I_{j_1}$, and hence

$$m_{id}(I) \leq m_{id}(I_{j_1}) \leq \sum_{k=1}^n m_{id}(I_{j_k}).$$

However, if $b \geq b_1$, then there exists another open interval, say, $I_{j_2} = (a_2, b_2)$ such that $a_2 < b_1 < b_2$. This yields a finite number of open intervals, say, $\{I_{j_k} = (a_k, b_k)\}_{k=1}^l$ with $a_1 < a$ and $a_k < b_{k-1} < b_k$ as well as $b < b_l$. Accordingly,

$$\sum_{k=1}^n m_{id}(I_{j_k}) \geq \sum_{k=1}^l (b_k - a_k) = b_l + \sum_{k=2}^l (b_{k-1} - a_k) - a_1 > b - a = m_{id}(I).$$

So, $m_{id,*}(I) = m_{id}(I)$.

The second is that I is a bounded interval. Regarding this case, for a sufficiently small $\varepsilon > 0$ there exists a compact interval $J \subseteq I$ such that $m_{id}(I) < m_{id}(J) + \varepsilon$. Consequently,

$$m_{id,*}(I) \geq m_{id,*}(J) = m_{id}(J) > m_{id}(I) - \varepsilon.$$

Thus, the desired estimate

$$m_{id,*}(I) \geq m_{id}(I) \geq m_{id,*}(I)$$

follows.

The third is that I is an unbounded interval. Regarding this case, for any $\ell > 0$ there exists a compact interval $J \subseteq I$ such that $m_{id,*}(J) = m_{id}(J) = \ell$. Accordingly,

$$m_{id,*}(I) \geq m_{id,*}(J) = \ell \quad \text{and so} \quad m_{id,*}(I) = \infty = m_{id}(I).$$

(iv) First of all, according to (i) we may assume that for a Vitali's covering \mathcal{V} of E there is an open set O with $m_{id}(O) < \infty$ and $E \subseteq O$. Note that this last inclusion implies that if $x \in E$ then there are an open interval $(x-r, x+r) \subseteq O$ and an interval $I = [a, b] \in \mathcal{V}$ with $m_{id}(I) < 2^{-1}r$. So, it follows that

$$\begin{cases} b < a + 2^{-1}r; \\ a \leq x \leq b; \\ y \in I \Rightarrow |x-y| \leq |x-a| + |y-a| < r; \\ y \in (x-r, x+r) \subseteq O. \end{cases}$$

Accordingly, we may further assume that any interval $I \in \mathcal{V}$ is a subset of O , thereby getting that if $\{I_j\}_{j=1}^\infty \subseteq \mathcal{V}$ are mutually disjoint then

$$\sum_{j=1}^\infty m_{id}(I_j) \leq m_{id}(O) < \infty.$$

Moreover, because I_j is closed and \mathcal{V} is a Vitali's covering of E , it follows that

$$\{I_j\}_{j=1}^n \subseteq \mathcal{V} \Rightarrow E \setminus \cup_{j=1}^n I_j \subseteq \cup_{I \in \mathcal{V}_n} I$$

where

$$\mathcal{V}_n = \{I \in \mathcal{V} : I \cap \cup_{j=1}^n I_j = \emptyset\} \text{ for all } n \in \mathbb{N}.$$

As a matter of fact, if $x \in E \setminus \cup_{j=1}^n I_j$, then $x \notin I_j$ for any $j \in \{1, \dots, n\}$, and the closedness of I_j produces $(x-r_j, x+r_j) \cap I_j = \emptyset$. Upon choosing $r_0 = 2^{-1} \min\{r_1, \dots, r_n\}$, we get a closed interval $I_{r_0} \in \mathcal{V}$ such that $m_{id}(I_{r_0}) < r < 2^{-1}r_j$ and $x \in I_{r_0}$, thereby finding

$$\begin{cases} I_{r_0} \subseteq (x-r_j, x+r_j); \\ I_{r_0} \cap I_j = \emptyset; \\ x \in I_{r_0} \subseteq \cup_{I \in \mathcal{V}_n} I. \end{cases}$$

If there exists a finite mutually disjoint subfamily of \mathcal{V} which covers E , then the argument is over. Otherwise, we can inductively select a mutually disjoint countable subfamily $\{I_j = [c_j - r_j, c_j + r_j]\}_{j=1}^\infty$ of \mathcal{V} such that

$$J_k = [c_k - 5r_k, c_k + 5r_k] \Rightarrow E \setminus \cup_{j=1}^n I_j \subseteq \cup_{j=n+1}^\infty J_j \text{ for all } k, n \in \mathbb{N}.$$

To be more precisely, let $n \in \mathbf{N}$ and $I_1 \in \mathcal{V}$ be selected arbitrarily, and suppose that the mutually disjoint closed intervals $\{I_j\}_{j=1}^n \subseteq \mathcal{V}$ have been chosen. Since $E \not\subseteq \cup_{j=1}^n I_j$, it follows that

$$\mathcal{V}'_n \neq \emptyset \text{ and so } \kappa_n = \sup \{m_{id}(I) : I \in \mathcal{V}'_n\} \leq m_{id}(O) < \infty.$$

Now, via choosing

$$I_{n+1} \in \mathcal{V}'_{n+1} \text{ with } m_{id}(I_{n+1}) > 2^{-1}\kappa_n,$$

we obtain inductively a countable mutually disjoint subfamily of \mathcal{V} such that

$$I \in \mathcal{V} \text{ and } I \cap \cup_{j=1}^n I_j = \emptyset \Rightarrow m_{id}(I_{n+1}) > 2^{-1}m_{id}(I).$$

Since

$$\sum_{j=1}^{\infty} m_{id}(I_j) < \infty \Rightarrow \lim_{j \rightarrow \infty} m_{id}(I_j) = 0,$$

in order to validate

$$E \setminus \cup_{j=1}^n I_j \subseteq \cup_{j=n+1}^{\infty} J_j$$

let $x \in E \setminus \cup_{j=1}^n I_j$. Then $x \in \cup_{I \in \mathcal{V}'_n} I$ and hence there exists $I \in \mathcal{V}$ which is disjoint from $\cup_{j=1}^n I_j$ and contains x . Now, this closed interval I must enjoy the property $I \cap I_{j_0} \neq \emptyset$ for some $j_0 \in \{1, \dots, n\}$ - otherwise - $m_{id}(I_k) > 2^{-1}m_{id}(I)$ for all $k \in \mathbf{N}$, contradicting $\lim_{k \rightarrow \infty} m_{id}(I_k) = 0$. Suppose that n_0 is the first integer obeying $x \in I \cap I_{n_0} \neq \emptyset$. Then $n_0 > n$. Thanks to $I \cap \cup_{j=1}^{n_0-1} I_j = \emptyset$, we have $m_{id}(I_{n_0}) > 2^{-1}m_{id}(I)$. Note that the distance from x to the centre c_{n_0} of I_{n_0} is at most $m_{id}(I) + 2^{-1}m_{id}(I_{n_0})$. Thus $m_{id}(I) < 2m_{id}(I_{n_0})$. This in turn implies that the distance from x to c_{n_0} is less than $2^{-1}5m_{id}(I_{n_0})$ and hence

$$x \in J_{n_0} \subseteq \cup_{j=n+1} J_j \text{ due to } n+1 < n_0+1 \Rightarrow n+1 \leq n_0.$$

Now, for any $\varepsilon > 0$ there is $n \in \mathbf{N}$ such that $\sum_{j=n+1}^{\infty} m_{id}(I_j) < 5^{-1}\varepsilon$. Consequently, both (ii) and (iii) are used to derive

$$\begin{aligned} m_{id,*}(E \setminus \cup_{j=1}^n I_j) &\leq m_{id,*}(\cup_{j=n+1}^{\infty} J_j) \\ &\leq \sum_{j=n+1}^{\infty} m_{id,*}(J_j) \\ &= 5 \sum_{j=n+1}^{\infty} m_{id}(I_j) \\ &< \varepsilon. \end{aligned}$$

□

Remark 5.1.4. The property (ii-3) in Theorem 5.1.3 means that $m_{id,*}$ is countably subadditive. And yet, $m_{id,*}$ is not countably additive. To see this, we say that two points x, y in the closed interval $[0, 1]$ are equivalent, denoted $x \sim y$, whenever $x - y \in \mathbf{Q}$. According to the Axiom of Choice, we take such a set $E_* \subseteq [0, 1]$ that contains exactly one point from each equivalence class decided by \sim , thereby finding

(i)

$$r_1, r_2 \in \mathbf{Q} \text{ and } r_1 \neq r_2 \Rightarrow (E_* + r_1) \cap (E_* + r_2) = \emptyset;$$

(ii)

$$[0, 1] \subseteq \bigcup_{r \in \mathbf{Q} \cap [0, 1]} (E_* + r) \subseteq [-1, 2];$$

(iii)

$$1 = m_{id,*}([0, 1]) \leq m_{id,*}\left(\bigcup_{r \in \mathbf{Q} \cap [0, 1]} (E_* + r)\right) \leq m_{id,*}([-1, 2]) = 3;$$

(iv)

$$r \in \mathbf{Q} \cap [0, 1] \Rightarrow m_{id,*}(E_* + r) = m_{id,*}(E_*).$$

Accordingly, if $m_{id,*}$ is countably additive, then

$$\begin{aligned} 1 &\leq m_{id,*}\left(\bigcup_{r \in \mathbf{Q} \cap [0, 1]} (E_* + r)\right) \\ &= \sum_{r \in \mathbf{Q} \cap [0, 1]} m_{id,*}(E_* + r) \\ &= \sum_{r \in \mathbf{Q} \cap [0, 1]} m_{id,*}(E_*) \\ &\leq 3, \end{aligned}$$

and hence we have a contradiction no matter if $m_{id,*}(E_*)$ is finite or infinite. But nevertheless, E_* is not m_{id} -measurable - otherwise - we use (iii) of Theorem 5.1.3 to achieve $m_{id,*}(E_*) = m_{id}(E_*) < \infty$, which leads to the last contradiction.

5.2 Derivatives of Increasing Functions

Recall that if $f : [a, b] \rightarrow \mathbf{R}$ has derivative $f'(c)$ at any $c \in (a, b)$ then for any $\varepsilon > 0$ there is a closed interval $I_{c,\varepsilon}$ such that it contains c in its interior and enjoys

$$m_{id}(I_{c,\varepsilon}) < \varepsilon \text{ and } \left| \frac{f(x) - f(c)}{x - c} - f'(c) \right| < \varepsilon \text{ for } x \in I_{c,\varepsilon} \cap [a, b].$$

Then $\mathcal{V} = \{I_{c,\varepsilon}\}$ is a Vitali's covering of (a, b) . Accordingly, it is natural to consider four one-sided derivatives.

Definition 5.2.1. Given $f : [a, b] \rightarrow \mathbf{R}$ and $c \in (a, b)$.

(i) *Four one-sided limits:*

$$\begin{cases} \limsup_{0 < x-c \rightarrow 0} f(x) = \inf_{\varepsilon > 0} \sup_{0 < x-c < \varepsilon} f(x); \\ \liminf_{0 < x-c \rightarrow 0} f(x) = \sup_{\varepsilon > 0} \inf_{0 < x-c < \varepsilon} f(x); \\ \limsup_{0 > x-c \rightarrow 0} f(x) = \inf_{\varepsilon > 0} \sup_{0 < c-x < \varepsilon} f(x); \\ \liminf_{0 > x-c \rightarrow 0} f(x) = \sup_{\varepsilon > 0} \inf_{0 < c-x < \varepsilon} f(x), \end{cases}$$

which enjoys

$$\liminf_{0 < x-c \rightarrow 0} f(x) \leq \limsup_{0 < x-c \rightarrow 0} f(x) \quad \text{and} \quad \liminf_{0 > x-c \rightarrow 0} f(x) \leq \limsup_{0 > x-c \rightarrow 0} f(x).$$

Moreover, $\lim_{x \rightarrow c} f(x) = L \in \mathbf{R}$ exists if and only if

$$\liminf_{0 < x-c \rightarrow 0} f(x) = \limsup_{0 < x-c \rightarrow 0} f(x) = \liminf_{0 > x-c \rightarrow 0} f(x) = \limsup_{0 > x-c \rightarrow 0} f(x) = L.$$

(ii) *Four one-sided derivatives:*

$$\begin{cases} \overline{D^+}f(c) = \limsup_{0 < x-c \rightarrow 0} \frac{f(x)-f(c)}{x-c}; \\ \underline{D^+}f(c) = \liminf_{0 < x-c \rightarrow 0} \frac{f(x)-f(c)}{x-c}; \\ \overline{D^-}f(c) = \limsup_{0 > x-c \rightarrow 0} \frac{f(x)-f(c)}{x-c}; \\ \underline{D^-}f(c) = \liminf_{0 > x-c \rightarrow 0} \frac{f(x)-f(c)}{x-c}, \end{cases}$$

which enjoys

$$\underline{D^+}f(c) \leq \overline{D^+}f(c) \quad \text{and} \quad \underline{D^-}f(c) \leq \overline{D^-}f(c).$$

Moreover, $f'(c) = Q \in \mathbf{R}$ exists if and only if

$$\underline{D^+}f(c) = \overline{D^+}f(c) = \underline{D^-}f(c) = \overline{D^-}f(c) = Q.$$

Example 5.2.2. If

$$f(x) = \begin{cases} 2^{-2}|x|, & x \in \mathbf{Q} \\ 2^{-1}|x|, & x \in \mathbf{R} \setminus \mathbf{Q} \end{cases},$$

then

$$\begin{cases} \overline{D^+}f(0) = 2^{-1}; \underline{D^+}f(0) = 2^{-2}; \overline{D^-}f(0) = -2^{-2}; \underline{D^-}f(0) = -2^{-1}, \\ \overline{D^+}f(1) = \infty; \underline{D^+}f(1) = 2^{-2}; \overline{D^-}f(1) = 2^{-2}; \underline{D^-}f(1) = -\infty. \end{cases}$$

Proposition 5.2.3. *If $f : [a, b] \rightarrow \mathbf{R}$ is increasing and $-\infty < r < R < \infty$, then*

$$E = \{x \in (a, b) : \underline{D}^-f(x) < r < R < \overline{D}^+f(x)\}$$

satisfies $m_{id}(E) = 0$.

Proof. According to Theorem 5.1.3(i), for $\varepsilon > 0$ there is an open set $O \supseteq E$ such that

$$m_{id}(O) < m_{id,*}(E) + \varepsilon.$$

On the one hand, if $x \in E$, then $\underline{D}^-f(x) < r$ ensures a sufficiently small $\delta \in (0, \varepsilon)$ such that

$$\delta^{-1}(f(x) - f(x - \delta)) < r \text{ and } [x - \delta, x] \subseteq O,$$

and hence the induced intervals $[x - \delta, x]$ form a Vitali's covering of E . Now, Theorem 5.1.3(iv) is used to produce mutually disjoint intervals

$$\{I_j = [x_j - \delta_j, x_j]\}_{j=1}^{n_0} \text{ with } m_{id,*}(E \setminus \cup_{j=1}^{n_0} I_j) < \varepsilon.$$

Upon setting $F = E \cap (\cup_{j=1}^{n_0} I_j)$, we have

$$\begin{cases} m_{id,*}(F) > m_{id,*}(E) - \varepsilon; \\ \sum_{j=1}^{n_0} (f(x_j) - f(x_j - \delta_j)) < r \sum_{j=1}^{n_0} \delta_j < r m_{id}(O) < r(m_{id,*}(E) + \varepsilon). \end{cases}$$

On the other hand, if $y \in F \subseteq E$, then $R < \overline{D}^+f(y)$, and hence there are a sufficiently small $\eta > 0$ and a $k \in \{1, \dots, n_0\}$ such that

$$[y, y + \eta] \subseteq I_k \text{ and } R\eta < f(y + \eta) - f(y).$$

Accordingly, the induced intervals $[y, y + \eta]$ form a Vitali's covering of F . Again, Theorem 5.1.3(iv) secures mutually disjoint intervals

$$\{J_j = [y_j, y_j + \eta_j]\}_{j=1}^{m_0} \text{ with } m_{id,*}(F \setminus \cup_{j=1}^{m_0} J_j) < \varepsilon.$$

However, since

$$F = (F \setminus \cap_{j=1}^{m_0} J_j) \cup (F \cap (\cup_{j=1}^{m_0} J_j)),$$

Theorem 5.1.3(ii-3) derives

$$m_{id,*}(F) \leq m_{id,*}(F \setminus \cap_{j=1}^{m_0} J_j) + m_{id,*}(F \cap (\cup_{j=1}^{m_0} J_j)).$$

Consequently, (ii-2), (ii-3) and (iii) of Theorem 5.1.3 give

$$\begin{aligned}
 \sum_{j=1}^{m_0} \eta_j &= \sum_{j=1}^{m_0} m_{id,*}(J_j) \\
 &\geq \sum_{j=1}^{m_0} m_{id,*}(F \cap J_j) \\
 &\geq m_{id,*}(F \cap (\cup_{j=1}^{m_0} J_j)) \\
 &> m_{id,*}(E) - 2\varepsilon.
 \end{aligned}$$

This in turn implies

$$\sum_{j=1}^{m_0} (f(y_j + \eta_j) - f(y_j)) > R \sum_{j=1}^{m_0} \eta_j > R(m_{id,*}(E) - 2\varepsilon).$$

At the same time, noticing that each J_j is a subset of some I_n , we sum only over those j 's with $J_j \subseteq I_n$ and use the hypothesis that f is increasing to obtain

$$\sum_{j: J_j \subseteq I_n} (f(y_j + \eta_j) - f(y_j)) \leq f(x_n) - f(x_n - h_n).$$

Furthermore, via summing over all such n , we estimate

$$\sum_{n=1}^{n_0} (f(x_n) - f(x_n - h_n)) \geq \sum_{j=1}^{m_0} (f(y_j + \eta_j) - f(y_j)) > R(m_{id,*}(E) - 2\varepsilon).$$

The above two-fold argument, plus $r < R$, derives

$$R(m_{id,*}(E) - 2\varepsilon) < r(m_{id,*}(E) + \varepsilon) \quad \text{and} \quad m_{id,*}(E) < (R - r)^{-1}(r + 2R)\varepsilon.$$

Therefore, $m_{id}(E) = 0$. □

Below is Lebesgue's theorem on the derivative of an increasing function.

Theorem 5.2.4. *If $f : [a, b] \rightarrow \mathbf{R}$ is increasing, then f' exists m_{id} -a.e. on $[a, b]$, is nonnegative and enjoys*

$$\int_{[a,b]} f' dm_{id} \leq f(b) - f(a)$$

with the strict inequality holding for any increasing step function f with $f(a) < f(b)$.

Proof. Note that

$$\{x \in (a, b) : \underline{D}^-f(x) < \overline{D}^+f(x)\}$$

is equal to

$$\bigcup_{r, R \in \mathbf{Q}} \{x \in (a, b) : \underline{D}^-f(x) < r < R < \overline{D}^+f(x)\}.$$

So, Proposition 5.2.3 deduces

$$m_{id}(\{x \in (a, b) : \underline{D}^-f(x) < \overline{D}^+f(x)\}) = 0,$$

whence

$$\underline{D}^-f(x) \geq \overline{D}^+f(x) \text{ for } m_{id} - \text{a.e. } x \in [a, b].$$

Since $-f(-x)$ is increasing on $[-b, -a]$, substituting $-f(-x)$ for $f(x)$ in the last a.e.-statement gives that

$$\overline{D}^+f(x) \leq \underline{D}^-f(x) \text{ for } m_{id} - \text{a.e. } x \in [a, b].$$

Consequently,

$$\overline{D}^+f(x) \leq \underline{D}^-f(x) \leq \overline{D}^-f(x) \leq \underline{D}^+f(x) \text{ for } m_{id} - \text{a.e. } x \in [a, b].$$

This in turn implies the existence of $f'(x)$ for m_{id} -a.e. $x \in [a, b]$.

In order to verify the desired integral inequality, for $n \in \mathbf{N}$ we extend the definition of f via setting $f(x) = f(b)$ as $x > b$ and define m_{id} -measurable function

$$f_n(x) = n(f(x + n^{-1}) - f(x)) \text{ for all } n \in \mathbf{N}.$$

Clearly, $f_n(x) \geq 0$ and $f_n(x) \rightarrow f'(x) \geq 0$ for m_{id} -a.e. $x \in [a, b]$. Since f' is m_{id} -measurable, we can utilize Fatou's lemma and the monotonicity of f to estimate

$$\begin{aligned} \int_{[a, b]} f' dm_{id} &\leq \liminf_{n \rightarrow \infty} \int_{[a, b]} f_n dm_{id} \\ &= n \liminf_{n \rightarrow \infty} \int_{[a, b]} (f(x + n^{-1}) - f(x)) dm_{id}(x) \\ &= n \liminf_{n \rightarrow \infty} \left(\int_{a+n^{-1}}^{b+n^{-1}} f dm_{id} - \int_{[a, b]} f dm_{id} \right) \\ &= n \liminf_{n \rightarrow \infty} \left(\int_b^{b+n^{-1}} f dm_{id} - \int_a^{a+n^{-1}} f dm_{id} \right) \\ &= f(b) - n \limsup_{n \rightarrow \infty} \int_a^{a+n^{-1}} f dm_{id} \\ &\leq f(b) - f(a). \end{aligned}$$

Note that if f is an increasing step function on $[a, b]$ then $f' = 0$ m_{id} -a.e. on $[a, b]$. So if $f(a) < f(b)$ is further assumed, then there is a strict inequality

$$\int_{[a,b]} f' dm_{id} = 0 < f(b) - f(a).$$

□

5.3 Absolutely Continuous Functions

The class $BV[a, b]$ more or less leads to the following concept.

Definition 5.3.1. A function $f: [a, b] \rightarrow \mathbf{R}$ is said to be absolutely continuous on $[a, b]$ provided that for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any finite sequence of mutually disjoint intervals $(a_1, b_1), \dots, (a_n, b_n) \subseteq [a, b]$ one has the implication

$$\sum_{j=1}^n (b_j - a_j) < \delta \Rightarrow \sum_{j=1}^n |f(b_j) - f(a_j)| < \varepsilon.$$

The class of all absolutely continuous functions on $[a, b]$ is written as $AC[a, b]$.

Example 5.3.2. If

(i) $x \mapsto f(x) = x^{\frac{3}{2}} \sin(x^{-1})$ with $f(0) = 0$ is in $AC[a, b]$.

(ii) if $0 < \alpha \leq 2^{-1}$ then $x \mapsto f_{2^{-1}}(x) = \sqrt{x}$ is in $AC[0, 1]$ and satisfies

$$\sup_{x \neq y \text{ in } [0, 1]} \frac{|f_{2^{-1}}(x) - f_{2^{-1}}(y)|}{|x - y|^\alpha} < \infty.$$

Proposition 5.3.3. The basic properties of $AC[a, b]$ are listed below.

(i) The sum, difference and product of two $AC[a, b]$ -functions are also $AC[a, b]$ -functions.

(ii) If an $AC[a, b]$ -function is nowhere zero, then its reciprocal is also an $AC[a, b]$ -function.

(iii) If f is of Lipschitz class on $[a, b]$, denoted $f \in Lip[a, b]$ - namely -

$$\sup_{x \neq y \text{ in } [a, b]} \frac{|f(x) - f(y)|}{|x - y|} < \infty,$$

equivalently, for any $\varepsilon > 0$ there is a $\delta > 0$ such that for any finite family of open intervals $(x_j, y_j) \subseteq [a, b]$ with

$$\sum_{j=1}^n (y_j - x_j) < \delta \Rightarrow \sum_{j=1}^n |f(x_j) - f(y_j)| < \varepsilon,$$

then $f \in AC[a, b]$, and hence f is uniformly continuous on $[a, b]$, denoted $f \in UC[a, b]$ - i.e. - for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$x, y \in [a, b] \text{ and } |x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon.$$

(iv) If $f \in Lip[c, d]$ and $\phi \in AC[a, b]$ obeys $\phi([a, b]) \subseteq [c, d] \subseteq \mathbf{R}$, then $f \circ \phi \in AC[a, b]$.

(v) If $f \in AC[a, b]$, then $f \in BV[a, b]$, and hence f is differentiable m_{id} -a.e. on $[a, b]$ with

$$\int_{[a, b]} |f'| dm_{id} \leq 2V_a^b f - (f(b) - f(a)).$$

(vi) Any $f \in AC[a, b]$ has Luzin's invariant property -namely- if $E \subseteq [a, b]$ satisfies $m_{id}(E) = 0$ then $f(E) \subseteq \mathbf{R}$ satisfies $m_{id}(f(E)) = 0$.

Proof. Since (i)-(ii)-(iii) follow directly from Definition 5.3.1, we are about to validate (iv)-(v)-(vi).

Firstly, for (iv), we use the condition that f is Lipschitz continuous on $[c, d]$ to get a constant

$$M = 1 + \sup_{x \neq y \text{ in } [c, d]} \frac{|f(x) - f(y)|}{|x - y|} < \infty$$

such that

$$|f(x) - f(y)| \leq M|x - y| \text{ for all } x, y \in [c, d].$$

Also, $\phi \in AC[a, b]$ implies that given $\varepsilon > 0$ there is $\delta > 0$ such that

$$\sum_{j=1}^n |\phi(x_j) - \phi(x_{j-1})| < M^{-1}\varepsilon$$

holds for any partition

$$\{a = x_0 < x_1 < \dots < x_n = b\} \text{ with } \sum_{j=1}^n (x_j - x_{j-1}) < \delta.$$

Accordingly,

$$\sum_{j=1}^n |f \circ \phi(x_j) - f \circ \phi(x_{j-1})| \leq M \sum_{j=1}^n |\phi(x_j) - \phi(x_{j-1})| < \varepsilon,$$

namely, $f \circ \phi \in AC[a, b]$.

Secondly, in order to verify (v), suppose $f \in AC[a, b]$. Definition 5.3.1 ensures a $\delta > 0$ such that for any finite family of mutually disjoint open intervals $(x_1, y_1), \dots, (x_n, y_n) \subseteq [a, b]$ there is the implication

$$\sum_{j=1}^n (y_j - x_j) < \delta \Rightarrow \sum_{j=1}^n |f(y_j) - f(x_j)| < 1.$$

Consequently, if the natural number n_0 is bigger than $(b-a)\delta^{-1}$, then

$$a_j = a + j(b-a)n_0^{-1} \text{ and } j \in \{0, 1, \dots, n_0\} \Rightarrow a_j - a_{j-1} < \delta \text{ and } V_{a_{j-1}}^{a_j} f \leq 1,$$

and hence $f \in BV[a, b]$ follows from

$$V_a^b f = \sum_{j=1}^{n_0} V_{a_{j-1}}^{a_j} f \leq n_0.$$

According to Theorem 3.1.7, there are two increasing functions

$$f_1(x) = V_{[a,x]} f \text{ and } f_2(x) = V_{[a,x]} f - f(x)$$

such that $f = f_1 - f_2$. This decomposition, along with Theorem 5.2.4, implies that f is differentiable m_{id} -a.e. on $[a, b]$ and enjoys

$$\begin{aligned} \int_a^b |f'| dm_{id} &\leq \int_a^b f'_1 dm_{id} + \int_a^b f'_2 dm_{id} \\ &\leq f_1(b) - f_1(a) + f_2(b) - f_2(a) \\ &= V_a^b f + V_a^b f - (f(b) - f(a)) \\ &= 2V_a^b f - (f(b) - f(a)) \end{aligned}$$

Thirdly, we validate (vi). To do so, let $f \in AC[a, b]$ and $E \subseteq [a, b]$ obey $m_{id}(E) = 0$. Then for any $\varepsilon > 0$ there are a $\delta > 0$ and a sequence of mutually disjoint open intervals $\{(a_j, b_j)\}_{j=1}^\infty$ such that

$$\begin{cases} E \subseteq \cup_{j=1}^\infty (a_j, b_j); \\ \sum_{j=1}^\infty (b_j - a_j) < \delta; \\ \sum_{j=1}^\infty |f(b_j) - f(a_j)| < \varepsilon. \end{cases}$$

According to (iii), $f \in C[a, b]$ and consequently, there are $x_j, y_j \in [a_j, b_j]$ such that

$$f(x_j) = \min_{z \in [a_j, b_j]} f(z) = m_j \quad \text{and} \quad f(y_j) = \max_{z \in [a_j, b_j]} f(z) = M_j.$$

This, along with $f \in AC[a, b]$, implies

$$\begin{cases} f(E) \subseteq \bigcup_{j=1}^{\infty} f([a_j, b_j]) \subseteq \bigcup_{j=1}^{\infty} [m_j, M_j]; \\ \sum_{j=1}^{\infty} |x_j - y_j| \leq \sum_{j=1}^{\infty} (b_j - a_j) < \delta; \\ \sum_{j=1}^{\infty} (M_j - m_j) = \sum_{j=1}^{\infty} |f(y_j) - f(x_j)| < \varepsilon. \end{cases}$$

In other words, $f(E)$ enjoys $m_{id}(f(E)) = 0$. □

The above assertion leads to two alternatives of $AC[a, b]$.

Theorem 5.3.4. *The following statements on a function $f : [a, b] \rightarrow \mathbf{R}$ are equivalent:*

- (i) $f \in AC[a, b]$;
- (ii) f has a derivative $f' \in LRS_{id}[a, b]$ and

$$f(x) = f(a) + \int_{[a, x]} f' dm_{id} \quad \text{for all } x \in [a, b];$$

- (iii) $f \in C[a, b] \cap BV[a, b]$ and $m_{id}(E) = 0 \Rightarrow m_{id}(f(E)) = 0$.

Here, (i) \Leftrightarrow (ii) and (i) \Leftrightarrow (iii) are known as the fundamental theorem of Lebesgue integration and the Banach-Zarecki theorem (due to Stefan Banach and Moisej Abramovitch Zarecki) respectively.

Proof. The result will be verified according to (i) \Leftrightarrow (ii) and (i) \Leftrightarrow (iii).

(i) \Leftrightarrow (ii) Suppose that (i) holds. Proposition 5.3.3(v) implies that $f \in BV[a, b]$ and so f' exists m_{id} -a.e on $[a, b]$ and m_{id} -integrable. If

$$\phi(x) = f(a) + \int_{[a, x]} f' dm_{id} \quad \text{for all } x \in [a, b],$$

then

$$\phi \in AC[a, b] \quad \text{and} \quad (\phi - f)'(x) = 0 \quad \text{for } m_{id} - \text{a.e } x \in [a, b].$$

Next, we are about to prove the desired formula $\phi = f$ on $[a, b]$. To do so, let $c \in (a, b)$. Then $\phi - f \in AC[a, c]$ and there is an m_{id} -measurable set $E \subseteq (a, c)$

with $m_{id}(E) = c - a$ such that $\phi' = f'$ on E . Let $\varepsilon > 0$ and $\delta > 0$ be chosen as by the definition of $\phi - f \in AC[a, c]$. If $x \in E$, then there is $h_x > 0$ such that

$$|(\phi - f)(x + h) - (\phi - f)(x)| < \varepsilon h \text{ for all } h \in (0, h_x),$$

and hence

$$\mathcal{V} = \{[x, x + h] : h \in (0, h_x) \text{ and } x \in E\}$$

is a Vitali's covering of E . Now, $m_{id}(E) = c - a$ and Theorem 5.1.3(iv) produce mutually disjoint intervals $\{[x_j, x_j + h_j]\}_{j=1}^n$ from \mathcal{V} such that

$$\cup_{j=1}^n (x_j, x_j + h_j) \subseteq (a, c) \text{ and } m_{id}(E \setminus \cup_{j=1}^n [x_j, x_j + h_j]) < \varepsilon.$$

Upon arranging the intervals $\{[x_j, x_j + h_j]\}_{j=1}^n$ so that $x_j + h_j < x_{j+1}$, setting $a = x_0; h_0 = 0; c = x_{n+1}$, and using $\phi - f \in AC[a, c]$, we obtain

$$\begin{cases} \sum_{j=0}^n |x_{j+1} - (x_j + h_j)| < \delta; \\ \sum_{j=0}^n |(\phi - f)(x_{j+1}) - (\phi - f)(x_j + h_j)| < \varepsilon; \\ \sum_{j=0}^n |(\phi - f)(x_j + h_j) - (\phi - f)(x_j)| < \varepsilon \sum_{j=0}^n h_j \leq \varepsilon(c - a), \end{cases}$$

whence

$$\begin{aligned} |(\phi - f)(c) - (\phi - f)(a)| &= \left| \sum_{j=0}^n (\phi - f)(x_{j+1}) - (\phi - f)(x_j) \right| \\ &\leq \sum_{j=0}^n |(\phi - f)(x_{j+1}) - (\phi - f)(x_j + h_j)| \\ &\quad + \sum_{j=0}^n |(\phi - f)(x_j + h_j) - (\phi - f)(x_j)| \\ &\leq \varepsilon(1 + c - a). \end{aligned}$$

Letting $\varepsilon \rightarrow 0$ gives $(\phi - f)(c) = (\phi - f)(a) = 0$, whence $\phi = f$ on $[a, b]$.

Conversely, suppose that (ii) is valid. Then there exists a function $g = f' \in LRS_{id}[a, b]$ satisfying

$$f(x) = f(a) + \int_{[a, x]} g dm_{id} \text{ for all } x \in [a, b],$$

and thus it is enough to verify that $x \mapsto \int_{[a, x]} g dm_{id}$ is in $AC[a, b]$. To do so, for $\varepsilon > 0$ and $n \in \mathbb{N}$, let $g_n = \min\{|g|, n\}$. Then the monotone convergence theorem

yields an $n_0 \in \mathbf{N}$ such that

$$\int_{[a,b]} (|g| - g_{n_0}) dm_{id} < 2^{-1}\epsilon.$$

Upon taking $\delta = (2n_0)^{-1}\epsilon$, we obtain that if $m_{id}(E) < \delta$ then

$$\int_E |g| dm_{id} \leq \int_{[a,b]} (|g| - g_{n_0}) dm_{id} + \int_E g_{n_0} dm_{id} < 2^{-1}\epsilon + n_0 m_{id}(E) < \epsilon.$$

Next, assume that $\{(a_j, b_j)\}_{j=1}^n$ is a finite family of mutually disjoint open subintervals of $[a, b]$ with $\sum_{j=1}^n (b_j - a_j) < \delta$. If $E = \cup_{j=1}^n (a_j, b_j)$, then

$$m_{id}(E) < \delta \text{ and } \int_E |g| dm_{id} < \epsilon,$$

and hence

$$\begin{aligned} \sum_{j=1}^n |f(b_j) - f(a_j)| &= \sum_{j=1}^n \left| \int_{(a_j, b_j)} g dm_{id} \right| \\ &\leq \sum_{j=1}^n \int_{(a_j, b_j)} |g| dm_{id} \\ &= \int_E |g| dm_{id} \\ &< \epsilon. \end{aligned}$$

Therefore, (i) holds.

(i) \Leftrightarrow (iii) Suppose that (i) is valid. Then (iii) follows from both (iii) and (v)-(vi) of Proposition 5.3.3.

Conversely, let (iii) hold. Since $f \in BV[a, b]$, Theorem 5.2.4 or Proposition 5.3.3(v) ensures $f' \in LRS_{id}[a, b]$. Consequently, for any $\epsilon > 0$ there is a step function $\psi : [a, b] \rightarrow \mathbf{R}$ such that

$$\int_{[a,b]} |f' - \psi| dm_{id} < 2^{-1}\epsilon.$$

Moreover, if $E_\epsilon \subseteq [a, b]$ is m_{id} -measurable with

$$m_{id}(E_\epsilon) < 2^{-1}\epsilon \left(1 + \sup_{x \in [a,b]} |\psi(x)| \right)^{-1},$$

then

$$\begin{aligned}
 \int_{E_\varepsilon} |f'| dm_{id} &\leq \int_{E_\varepsilon} |f' - \Psi| dm_{id} + \int_{E_\varepsilon} |\Psi| dm_{id} \\
 &\leq \int_{[a,b]} |f' - \Psi| dm_{id} + \left(\sup_{x \in [a,b]} |\Psi(x)| \right) m_{id}(E_\varepsilon) \\
 &< 2^{-1}\varepsilon + 2^{-1}\varepsilon \\
 &= \varepsilon.
 \end{aligned}$$

Also, f' exists m_{id} -a.e. on $[a, b]$ - in other words - if D stands for the points in $[a, b]$ where f' exists, then $m_{id}([a, b] \setminus D) = 0$. If $[c, d] \subseteq [a, b]$, then

$$f \in C[a, b] \Rightarrow [\min\{f(c), f(d)\}, \max\{f(c), f(d)\}] \subseteq f([c, d]),$$

and hence setting

$$E = [c, d] \cap D \text{ and } F = [c, d] \setminus D,$$

along with Theorem 5.1.3(ii-1) and the second condition in (iii), yields

$$\begin{cases} [c, d] = E \cup F; \\ m_{id}(F) = 0; \\ m_{id}(f(F)) = 0; \\ |f(d) - f(c)| \leq m_{id,*}(f([c, d])) \leq m_{id,*}(f(E)) + m_{id,*}(f(F)). \end{cases}$$

Next, let us control $m_{id,*}(f(E))$ from above.

Step 1 – proving that if $\varepsilon > 0$ and $k \in \mathbf{N}$ and

$$E_{\varepsilon,k} = \{x \in E : (k-1)\varepsilon \leq |f'(x)| < \varepsilon k\}$$

then

$$E_{\varepsilon,k} \cap E_{\varepsilon,l} = \emptyset \text{ for } k \neq l \text{ and } m_{id,*}(f(E_{\varepsilon,k})) \leq \varepsilon k m_{id,*}(E_{\varepsilon,k}).$$

In fact, note that for any $\eta > 0$ and $x \in E_{\varepsilon,k}$ there is $n_x \in \mathbf{N}$ enjoying the implication

$$y \in E_{\varepsilon,k} \text{ and } |x - y| < n_x^{-1} \Rightarrow |f(x) - f(y)| \leq (\varepsilon k + \eta)|x - y|.$$

So, for each $n \in \mathbf{N}$ let $E_{\varepsilon,k;\eta,n}$ be the set of all points $x \in E_{\varepsilon,k}$ with the implication

$$y \in E_{\varepsilon,k} \text{ and } |x - y| < n^{-1} \Rightarrow |f(x) - f(y)| \leq (\varepsilon k + \eta)|x - y|.$$

Then we use Theorem 5.1.3(ii) to get

$$\begin{cases} E_{\varepsilon,k;\eta,n} \subseteq E_{\varepsilon,k;\eta,n+1}; \\ E_{\varepsilon,k} = \bigcup_{n=1}^{\infty} E_{\varepsilon,k;\eta,n}; \\ m_{id,*}(E_{\varepsilon,k}) = \lim_{n \rightarrow \infty} m_{id,*}(E_{\varepsilon,k;\eta,n}); \\ f(E_{\varepsilon,k;\eta,n}) \subseteq f(E_{\varepsilon,k;\eta,n+1}); \\ f(E_{\varepsilon,k}) = \bigcup_{n=1}^{\infty} f(E_{\varepsilon,k;\eta,n}); \\ m_{id,*}(f(E_{\varepsilon,k})) = \lim_{n \rightarrow \infty} m_{id,*}(f(E_{\varepsilon,k;\eta,n})). \end{cases}$$

By Theorem 5.1.3(i) there is a sequence of intervals $\{I_{n,j}\}_{j=1}^{\infty}$ such that

$$\begin{cases} m_{id}(I_{n,j}) < n^{-1}; \\ E_{\varepsilon,k;\eta,n} \subseteq \bigcup_{j=1}^{\infty} I_{n,j}; \\ \sum_{j=1}^{\infty} m_{id}(I_{n,j}) \leq m_{id,*}(E_{\varepsilon,k;\eta,n}) + \eta. \end{cases}$$

Thus,

$$x, y \in E_{\varepsilon,k;\eta,n} \cap I_{n,j} \Rightarrow |x - y| < n^{-1} \Rightarrow |f(x) - f(y)| \leq (\varepsilon k + \eta)|x - y|.$$

As a consequence, we read off that $f(E_{\varepsilon,k;\eta,n} \cap I_{n,j})$ is a subset of an interval of length at most $(\varepsilon k + \eta)m_{id}(I_{n,j})$, thereby estimating

$$\begin{aligned} m_{id,*}(f(E_{\varepsilon,k;\eta,n})) &\leq \sum_{j=1}^{\infty} m_{id,*}(f(E_{\varepsilon,k;\eta,n} \cap I_{n,j})) \\ &\leq (\varepsilon k + \eta) \sum_{j=1}^{\infty} m_{id}(I_{n,j}) \\ &\leq (\varepsilon k + \eta)(m_{id,*}(E_{\varepsilon,k;\eta,n}) + \eta). \end{aligned}$$

This, along with Theorem 5.1.3(ii-2), derives

$$\begin{aligned} m_{id,*}(f(E_{\varepsilon,k})) &= \lim_{n \rightarrow \infty} m_{id,*}(f(E_{\varepsilon,k;\eta,n} \cap I_{n,j})) \\ &\leq (\varepsilon k + \eta) \lim_{n \rightarrow \infty} (m_{id,*}(E_{\varepsilon,k;\eta,n}) + \eta) \\ &= (\varepsilon k + \eta)(m_{id,*}(E_{\varepsilon,k}) + \eta). \end{aligned}$$

Upon sending η to 0 in the last estimation, we reach the desired inequality.

Step 2 – proving $m_{id,*}(f(E)) \leq \int_E |f'| dm_{id}$.

As a matter of fact, via using $E = \cup_{k=1}^{\infty} E_{\varepsilon,k}$ (which and whose $E_{\varepsilon,k}$ are m_{id} -measurable) and *Step 1* we obtain

$$\begin{aligned}
 m_{id,*}(f(E)) &= m_{id,*}(\cup_{k=1}^{\infty} f(E_{\varepsilon,k})) \\
 &\leq \sum_{k=1}^{\infty} m_{id,*}(f(E_{\varepsilon,k})) \\
 &\leq \sum_{k=1}^{\infty} (\varepsilon k) m_{id,*}(E_{\varepsilon,k}) \\
 &\leq \sum_{k=1}^{\infty} \varepsilon(k-1) m_{id,*}(E_{\varepsilon,k}) + \varepsilon \sum_{k=1}^{\infty} m_{id,*}(E_{\varepsilon,k}) \\
 &= \sum_{k=1}^{\infty} \varepsilon(k-1) m_{id}(E_{\varepsilon,k}) + \varepsilon \sum_{k=1}^{\infty} m_{id}(E_{\varepsilon,k}) \\
 &\leq \sum_{k=1}^{\infty} \int_{E_{\varepsilon,k}} |f'| dm_{id} + \varepsilon m_{id}(E),
 \end{aligned}$$

thereby arriving at the desired inequality due to $\varepsilon \rightarrow 0$.

Now, an application of *Step 2* derives

$$\begin{aligned}
 |f(d) - f(c)| &\leq m_{id,*}(f([c,d])) \\
 &\leq m_{id,*}(f(E)) + m_{id,*}(f(F)) \\
 &\leq \int_E |f'| dm_{id} + \int_F |f'| dm_{id} \\
 &= \int_{[c,d]} |f'| dm_{id}.
 \end{aligned}$$

Accordingly, for $\varepsilon > 0$ there is $\delta > 0$ such that if $\{(a_j, b_j)\}_{j=1}^n$ are mutually disjoint subintervals of $[a, b]$ with

$$m_{id}(\cup_{j=1}^n (a_j, b_j)) = \sum_{j=1}^n (b_j - a_j) < \delta$$

then

$$\int_{\cup_{j=1}^n [a_j, b_j]} |f'| dm_{id} < \varepsilon$$

and hence

$$\begin{aligned} \sum_{j=1}^n |f(b_j) - f(a_j)| &\leq \sum_{j=1}^n \int_{(a_j, b_j)} |f'| dm_{id} \\ &= \int_{\cup_{j=1}^n (a_j, b_j)} |f'| dm_{id} \\ &< \varepsilon, \end{aligned}$$

namely, $f \in AC[a, b]$.

□

As the first application of Theorem 5.3.4, we have the following Lebesgue's integration-by-part formula.

Corollary 5.3.5. *If $f, g \in AC[a, b]$, then*

$$\int_{[a, b]} f g' dm_{id} + \int_{[a, b]} f' g dm_{id} = f(b)g(b) - f(a)g(a).$$

Proof. Since $f, g \in AC[a, b]$, from Proposition 5.3.3(i) it follows that $fg \in AC[a, b]$. Consequently, Theorem 5.3.4 derives

$$\int_{[a, b]} (f'g + fg') dm_{id} = \int_{[a, b]} (fg)' dm_{id} = f(b)g(b) - f(a)g(a),$$

as desired.

□

As the second application of Theorem 5.3.4, we establish the following change-of-variable formula for $LRS_{id}[a, b]$.

Corollary 5.3.6. *If*

$$\left\{ \begin{array}{l} f \in LRS_{id}[c, d]; \\ \phi \in AC[a, b]; \\ \phi([a, b]) \subseteq [c, d]; \\ (f \circ \phi)\phi' \in LRS_{id}[a, b], \end{array} \right.$$

then

$$\int_{\phi([t_1, t_2])} f dm_{id} = \int_{[t_1, t_2]} (f \circ \phi)\phi' dm_{id} \text{ for all } a \leq t_1 < t_2 \leq b.$$

Proof. Upon setting $F(x) = \int_{[c,x]} f dm_{id}$, we obtain $F \in AC[c, d]$ and

$$(F \circ \phi)'(t) = f(\phi(t))\phi'(t) \text{ for } m_{id} - \text{a.e. } t \in [a, b],$$

thereby using Theorem 5.3.4 to derive that if $a \leq t_1 < t_2 \leq b$ then

$$\begin{aligned} \int_{\phi([t_1, t_2])} f dm_{id} &= \int_{\phi([t_1, t_2])} F' dm_{id} \\ &= (F \circ \phi)(t_2) - (F \circ \phi)(t_1) \\ &= \int_{[t_1, t_2]} (F \circ \phi)' dm_{id} \\ &= \int_{[t_1, t_2]} (f \circ \phi)\phi' dm_{id}, \end{aligned}$$

as desired. □

5.4 Cantor's Ternary Set and Singular Function

The Cantor ternary set C is constructed by iteratively deleting the open middle third from a family of line segments within the closed interval $[0, 1]$. More precisely, let

$$\begin{cases} C_0 = [0, 1]; \\ \alpha E + \beta = \{\alpha x + \beta : x \in E\} \text{ for } E \subseteq \mathbf{R} \text{ and } \alpha, \beta \in \mathbf{R}. \end{cases}$$

Then, the first step is to remove the open interval $(3^{-1}, 3^{-1} \cdot 2)$ from C_0 - the remaining part is

$$C_1 = [0, 3^{-1}] \cup [3^{-1} \cdot 2, 1].$$

We continue this process by removing the middle third of each interval in C_1 to reach the second step

$$C_2 = [0, 3^{-2}] \cup [3^{-2} \cdot 2, 3^{-1}] \cup [0, 3^{-1} \cdot 2] \cup [3^{-2} \cdot 2^3, 1].$$

For the $3 \leq n$ -th step, we remove the middle third of each interval in the recursively-defined C_{n-1} to get

$$C_n = 3^{-1}C_{n-1} \cup (3^{-1} \cdot 2 + 3^{-1}C_{n-1}).$$

Clearly, $\{C_k\}_{k=0}^{\infty}$ is a decreasing sequence of closed subsets of $[0, 1]$, and each $C_{k \geq 1}$ is the union of mutually disjoint 2^k closed intervals $\{M_{k,j}\}_{j=1}^{2^k}$ with $m_{id}(M_{k,j}) = 3^{-k}$. Now, Cantor's ternary set is defined by $C = \bigcap_{k=0}^{\infty} C_k$.

Theorem 5.4.1.

- (i) C is closed and yet has no isolated points.
- (ii) $m_{id}(C) = 0$.
- (iii) C consists of all ternary numbers $\sum_{j=1}^{\infty} 3^{-j} a_j$ with $a_j \in \{0, 2\}$.
- (iv) C is uncountable.

Proof. (i) Because each C_j is closed, so is $\bigcap_{j=0}^{\infty} C_j = C$. If $x \in C$ and $\varepsilon > 0$, then there is a big integer j such that $3^{-j} < \varepsilon$, and hence $x \in C_j \Rightarrow x \in [a, b]$ for some interval $M_{j,l} \subseteq C_j$ with $m_{id}(M_{j,l}) = 3^{-j}$. Accordingly, the endpoints of $M_{j,l}$ are in the interval $(x - \varepsilon, x + \varepsilon)$. Since the endpoints of intervals in C_j are not removed in subsequent $C_{k>j}$, there is a point $y \in C \cap ((x - \varepsilon, x + \varepsilon) \setminus \{x\})$. Thus, C contains no isolated points.

(ii) Note that $\{M_{k,j}\}_{k,j}$ are disjoint one from another. So it follows from Theorem 5.1.3(ii) that

$$\begin{aligned}
 m_{id,*}(C) &\leq m_{id,*}(C_k) \\
 &\leq m_{id,*}(\bigcup_{j=1}^{2^k} M_{k,j}) \\
 &\leq \sum_{j=1}^{2^k} m_{id,*}(M_{k,j}) \\
 &= (2 \cdot 3^{-1})^k \\
 &\rightarrow 0 \text{ as } k \rightarrow \infty.
 \end{aligned}$$

Consequently, $m_{id}(C) = 0$.

(iii) On the one hand, we have the following implication

$$\begin{aligned}
 x &= \sum_{j=1}^{\infty} 3^{-j} a_j \text{ with } a_j \in \{0, 2\} \\
 &\Rightarrow 3^{-1}x = \sum_{j=1}^{\infty} 3^{-j} b_j \text{ with } b_1 = 0 \text{ and } b_j = a_{j-1} \text{ for } j \geq 2 \\
 &\Rightarrow 3^{-1} \cdot 2 + 3^{-1}x = \sum_{j=1}^{\infty} 3^{-j} c_j \text{ with } c_1 = 2 \text{ and } c_j = a_{j-1} \text{ for } j \geq 2.
 \end{aligned}$$

Accordingly, $x \in C_1$ if and only if x is of form $3^{-1}y$ or $2 \cdot 3^{-1} + 3^{-1}y$ for some $y \in [0, 1]$. Upon repeating this argument, we find that $x \in C_k$ amounts to $x =$

$\sum_{j=1}^k 3^{-j}a_j$ with $a_j \in \{0, 2\}$ for $j \in \{1, \dots, k\}$. Since C is closed, $x = \sum_{j=1}^{\infty} 3^{-j}a_j$ with $a_j \in \{0, 2\}$ is in C .

On the other hand, let $x \in C$. If x 's ternary expansion $\sum_{j=1}^{\infty} 3^{-j}a_j$ with $a_j \in \{0, 2\}$ is unique, then the previous argument reveals $a_j \in \{0, 2\}$, and hence the desired result follows. Now it remains to verify the uniqueness. In fact, let x have two ternary expansions

$$\sum_{j=1}^{\infty} 3^{-j}a_j = x = \sum_{j=1}^{\infty} 3^{-j}b_j \quad \text{with } a_j, b_j \in \{0, 2\}.$$

If the statement " $a_j = b_j$ for all $j \in \mathbf{N}$ " is not valid, then there exists a minimal n such that $a_n \neq b_n$ - this ensures $a_j = b_j$ for all $j \in \{1, \dots, n-1\}$, and hence one of a_n or b_n is 0 and the other is 2. Without loss of generality, we may take $a_n = 0$ and $b_n = 2$, thereby getting

$$\begin{aligned} \sum_{j=1}^{n-1} 3^{-j}a_j + \sum_{j=n+1}^{\infty} 3^{-j}a_j &= \sum_{j=1}^{\infty} 3^{-j}a_j \\ &= \sum_{j=1}^{\infty} 3^{-j}b_j \\ &= \sum_{j=1}^{n-1} 3^{-j}b_j + 3^{-n} \cdot 2 + \sum_{j=n+1}^{\infty} 3^{-j}b_j. \end{aligned}$$

The choice of n gives $a_j = b_j$ for $j \in \{1, \dots, n-1\}$, and thus

$$3^{-n} \cdot 2 + \sum_{j=n+1}^{\infty} 3^{-j}b_j = \sum_{j=n+1}^{\infty} 3^{-j}a_j \leq 2 \sum_{j=n+1}^{\infty} 3^{-j} = 3^{-n}.$$

Consequently, we meet a contradiction

$$0 \leq \sum_{j=n+1}^{\infty} 3^{-j}b_j \leq -3^{-n} \quad \text{where } b_j \in \{0, 2\}.$$

So, $a_j = b_j$ for all $j \in \mathbf{N}$ - namely - the ternary representation of x is unique.

(iv) Since $C \subseteq [0, 1]$, it is enough to find a surjective map from C to $[0, 1]$ which is uncountable. To do so, according to (iii) we define

$$C \ni x = \sum_{j=1}^{\infty} 3^{-j}a_j \mapsto c(x) = \sum_{j=1}^{\infty} 2^{-1-j}a_j \in [0, 1].$$

Note that any $y \in [0, 1]$ has a binary expansion $y = \sum_{j=1}^{\infty} 2^{-j} b_j$ with $b_j \in \{0, 1\}$. So, if $x = \sum_{j=1}^{\infty} 3^{-j} (2b_j)$ then $c(x) = y$ - namely - c is surjective. \square

Next, we are about to extend the Cantor map c in Theorem 5.4.1 to $[0, 1]$. Let a and b be the left and right endpoints respectively of the same removed interval $(a, b) \in \{M_{k,j}\}$ which was removed in the k -th step. Then

$$\begin{cases} a = \sum_{j=1}^k 3^{-j} a_j + 0 \cdot 3^{-k} + \sum_{j=k+1}^{\infty} 2 \cdot 3^{-j}; \\ b = \sum_{j=1}^k 3^{-j} a_j + 2 \cdot 3^{-k} + \sum_{j=k+1}^{\infty} 0 \cdot 3^{-j}. \end{cases}$$

This yields

$$\begin{cases} c(a) = \sum_{j=1}^k 2^{-j-1} a_j + 0 \cdot 2^{-k} + \sum_{j=k+1}^{\infty} 1 \cdot 2^{-j} = \sum_{j=1}^k 2^{-1-j} a_j + 2^{-k}; \\ c(b) = \sum_{j=1}^k 2^{-1-j} a_j + 1 \cdot 2^{-k} + \sum_{j=k+1}^{\infty} 0 \cdot 2^{-j} = c(a). \end{cases}$$

Since the union of both C and the removed intervals is $[0, 1]$, Cantor's singular function $c_s : [0, 1] \rightarrow [0, 1]$ is now defined as

$$c_s(x) = \begin{cases} c(x) & , x \in C \\ \sup_{C \ni y \leq x} c(y) & , x \in [0, 1] \setminus C \end{cases}$$

and enjoys the following assertion.

Theorem 5.4.2.

- (i) c_s is increasing.
- (ii) c_s is continuous.
- (iii) c_s has zero derivative on $[0, 1] \setminus C$.
- (iv) c_s belongs to $BV[0, 1] \setminus AC[0, 1]$.

Proof. (i) Firstly, let

$$x = \sum_{j=1}^{\infty} 3^{-j} x_j, y = \sum_{j=1}^{\infty} 3^{-j} y_j \in C \text{ with } x_j, y_j \in \{0, 2\}.$$

If $x < y$, then x and y belong to the same interval until some k , and at that point y is in a later interval. Namely, for some k , $x_k < y_k$, for all $j < k$, $x_j = y_j$. However, $x_k < y_k$, plus $x_k, y_k \in \{0, 2\}$, implies $x_k = 0$ and $y_k = 2$. Accordingly,

$$c_s(x) = \sum_{j=1}^{k-1} 2^{-1-j} x_j + \sum_{j=k+1}^{\infty} 2^{-1-j} x_j \leq \sum_{j=1}^{k-1} 2^{-1-j} y_j + 2^{-k} \leq c_s(y).$$

Secondly, let $x, y \in [0, 1]$ with $x < y$ and $c_s(x) > c_s(y)$. There exist two points $x_0, y_0 \in \mathbb{C}$ obeying $c_s(x) = c_s(x_0)$ and $c_s(y) = c_s(y_0)$ - actually - x_0 and y_0 are chosen to be the left and right endpoints of the removed interval to which x and y belongs respectively. But, $c_s(x) \neq c_s(y)$ ensures that x_0 and y_0 must not correspond to the endpoints of the same interval. Meanwhile, $x < y$ implies $x_0 < y_0$. So

$$x_0, y_0 \in \mathbb{C} \Rightarrow c_s(x) = c_s(x_0) \leq c_s(y_0) = c_s(y) \text{ against } c_s(x) > c_s(y).$$

Consequently, c_s is increasing.

(ii) Evidently, c_s is continuous at $x \in [0, 1] \setminus \mathbb{C}$. Now, if $x \in \mathbb{C}$, then we consider two situations.

The first is to prove that if $y \in \mathbb{C}$ is close to x then $c_s(y)$ is close to $c_s(x)$. Without loss of generality, we may take $y > x$ and $y - x < 3^{-l}$ for any $l \in \mathbb{N}$. Upon taking $\{x_j\}$ and $\{y_j\}$ as the respective coefficients of x 's and y 's ternary expansions, we get

$$0 < y - x = \sum_{j=1}^{\infty} 3^{-j}(y_j - x_j) < 3^{-l},$$

thereby reaching $x_j = y_j$ for all $j < l$ - otherwise - if $j_0 < l$ is the minimal index enjoying $x_{j_0} \neq y_{j_0}$ then $x_{j_0} = 2, y_{j_0} = 0$ or $x_{j_0} = 0, y_{j_0} = 2$, and hence

$$y - x < 0 \text{ or } y - x > 3^{-l} \text{ contradicting } 0 < y - x < 3^{-l}.$$

Consequently, we achieve

$$c_s(y) - c_s(x) = \sum_{j=1}^{\infty} 2^{-1-j}(y_j - x_j) = \sum_{j=l}^{\infty} 2^{-1-j}(y_j - x_j) \leq 2^{-l},$$

thereby taking a sufficiently large l such that $2^{-l} < \varepsilon$ for any $\varepsilon > 0$.

The second is to prove that if $y \notin \mathbb{C}$ is close to x then $c_s(y)$ is also close to $c_s(x)$. In fact, for $y \notin \mathbb{C}$ there is $y_0 \in \mathbb{C}$ such that

$$c_s(y) = c_s(y_0) \text{ and } |x - y_0| \leq |x - y|,$$

via taking y_0 as the endpoint of the removed interval which y is in - there exist two such endpoints, so we take the one closer to x such that $|x - y_0| \leq |x - y|$. But then for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$|x - y_0| < \delta \Rightarrow |c_s(x) - c_s(y_0)| < \varepsilon.$$

Since $c_s(y) = c_s(y_0)$, we just take $|x - y| < \delta$.

(iii) If $x \in [0, 1] \setminus C$, then c_s is constant on an open interval around x , and hence

$$c'_s(x) = \lim_{t \rightarrow 0} t^{-1} (c_s(x+t) - c_s(x)) = 0.$$

(iv) According to the just-proved (iii), we have $c'_s(x) = 0$ for m_{id} -a.e. $x \in [0, 1]$ and

$$\int_{[0,1]} c'_s dm_{id} = 0 \neq c_s(1) - c_s(0),$$

thereby finding that the Fundamental Theorem of Lebesgue Calculus fails for c_s and so $c_s \notin AC[0, 1]$. However, $c_s \in BV[0, 1]$ due to the above-proved (i). \square

5.5 Lebesgue's Points

Needless to say, Cantor's singular function c_s is uniformly continuous on $[0, 1]$. However, there is a clear difference between the continuity and the uniform continuity of a function on \mathbf{R} .

Theorem 5.5.1. *Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be locally m_{id} -integrable - namely - $f1_{[-r,r]} \in LRS_{id}(\mathbf{R})$ for all $r > 0$.*

(i) *If f is continuous, denoted $f \in C(\mathbf{R})$, then*

$$\lim_{r \rightarrow 0} (2r)^{-1} \int_{(x-r, x+r)} |f - f(x)| dm_{id} = 0 \text{ for all } x \in \mathbf{R},$$

but not conversely.

(ii) *f is uniformly continuous on \mathbf{R} , denoted $f \in UC(\mathbf{R})$ - namely - for any $\varepsilon > 0$ there exists $\delta > 0$ such that $|y - x| < \delta \Rightarrow |f(y) - f(x)| < \varepsilon$ - if and only if - for any $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$(2r)^{-1} \int_{(x-r, x+r)} |f - f(x)| dm_{id} < \varepsilon \text{ for all } (x, r) \in \mathbf{R} \times (0, \delta).$$

Proof. (i) Suppose that $f \in C(\mathbf{R})$ and $x \in \mathbf{R}$. Then for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$|y - x| < r < \delta \Rightarrow |f(y) - f(x)| < \varepsilon.$$

This implies

$$(2r)^{-1} \int_{(x-r, x+r)} |f(y) - f(x)| dm_{id} \leq \varepsilon,$$

as desired.

For the converse, we may consider the following function $f_0 = \sum_{j=1}^{\infty} f_j$ where

$$f_j(x) = \begin{cases} j^3x - j^2 & , 0 \leq x - j^{-1} \leq 2^{-1}j^{-3} \\ 1 + j^2 - j^3x & , 2^{-1}j^{-3} \leq x - j^{-1} \leq j^{-3} \\ 0 & , x - j^{-1} \in \mathbf{R} \setminus [0, j^{-3}] \end{cases}.$$

Clearly, f_0 is discontinuous at 0, but if $(j_0 + 1)^{-1} \leq r < j_0^{-1}$ is valid for some $j_0 \in \mathbf{N}$, then a computation gives

$$\begin{aligned} (2r)^{-1} \int_{(-r,r)} |f_0 - f_0(0)| dm_{id} &= (2r)^{-1} \sum_{j=1}^{\infty} \int_{(0,r)} |f_j| dm_{id} \\ &\leq 2^{-1}(j_0 + 1) \sum_{j=j_0}^{\infty} \int_{(0,r)} |f_j| dm_{id} \\ &\leq 2^{-1}(j_0 + 1) \sum_{j=j_0}^{\infty} \int_{\mathbf{R}} |f_j| dm_{id} \\ &= 2^{-3}(j_0 + 1) \sum_{j=j_0}^{\infty} j^{-3} \\ &< (j_0 + 1)(j_0 - 1)^{-2} \\ &\rightarrow 0 \text{ as } j_0 \rightarrow \infty \text{ or } r \rightarrow 0. \end{aligned}$$

(ii) It is enough to verify the if-part. Assume that for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$(2r)^{-1} \int_{(x-r, x+r)} |f - f(x)| dm_{id} < \varepsilon \text{ for all } (x, r) \in \mathbf{R} \times (0, \delta).$$

If $f \notin UC(\mathbf{R})$, then there exists an $\varepsilon_0 > 0$ such that for every $\delta > 0$ there is a pair of points $\{x_0, y_0\}$ satisfying

$$|x_0 - y_0| < 2^{-1}\delta \text{ and } |f(x_0) - f(y_0)| > \varepsilon_0.$$

If $r_0 = 2|x_0 - y_0| < \delta$, then

$$(x_0 - 2^{-1}r_0, x_0 + 2^{-1}r_0) \subseteq (x_0 - r_0, x_0 + r_0) \cap (y_0 - r_0, y_0 + r_0),$$

and hence

$$\begin{aligned}
 & (2r_0)^{-1} \left(\int_{(x_0-r_0, x_0+r_0)} |f-f(x_0)| dm_{id} + \int_{(y_0-r, y_0+r)} |f-f(y_0)| dm_{id} \right) \\
 & \geq (2r_0)^{-1} \int_{(x_0-2^{-1}r, x_0+2^{-1}r)} (|f-f(x_0)| + |f-f(y_0)|) dm_{id} \\
 & \geq (2r_0)^{-1} \int_{(x_0-2^{-1}r, x_0+2^{-1}r)} |f(x_0) - f(y_0)| dm_{id} \\
 & \geq 2^{-1} \epsilon_0.
 \end{aligned}$$

Consequently, we have

$$2^{-2} \epsilon_0 \leq (2r_0)^{-1} \int_{(x_0-r_0, x_0+r_0)} |f-f(x_0)| dm_{id}$$

or

$$2^{-2} \epsilon_0 \leq (2r_0)^{-1} \int_{(y_0-r, y_0+r)} |f-f(y_0)| dm_{id}.$$

In other words, we get $\epsilon_1 = 2^{-2} \epsilon > 0$ such that for every $\delta > 0$ there are $z_0 \in \mathbf{R}$ and $r_0 \in (0, \delta)$ obeying

$$\epsilon_1 \leq (2r_0)^{-1} \int_{(z_0-r_0, z_0+r_0)} |f-f(z_0)| dm_{id},$$

contradicting the assumption. So, $f \in UC(\mathbf{R})$. □

The original Lebesgue's differentiation theorem states that if f is m_{id} -integrable on \mathbf{R} , i.e., $f \in LRS_{id}(\mathbf{R})$, then

$$F(x) = \int_{(-\infty, x)} f dm_{id}$$

has the following property

$$F'(x) = \lim_{r \rightarrow 0} (2r)^{-1} \int_{(x-\epsilon, x+\epsilon)} f dm_{id} = f(x) \text{ for } m_{id} - \text{a.e. } x \in \mathbf{R}.$$

Partially, this is an analogue or extension of Theorem 5.3.4 over a compact interval. Based on Theorem 5.5.1 there is a slightly stronger statement.

Theorem 5.5.2. *Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be locally m_{id} -integrable. Then*

$$\lim_{r \rightarrow 0} (2r)^{-1} \int_{(x-r, x+r)} |f - f(x)| dm_{id} = 0 \text{ for } m_{id} - a.e. \ x \in \mathbf{R}.$$

Accordingly, any point x for which this last formula is true is called *Lebesgue's point of f* .

Proof. Because x is Lebesgue's point of f if and only if x is Lebesgue's point of $f1_{(-r, r)}$ for any given $r > 0$, without loss of generality, we may assume $f \in LRS_{id}(\mathbf{R})$, and then verify the assertion according to the following two steps.

Step 1 - Hardy-Littlewood's maximal theorem: if $f \in LRS_{id}(\mathbf{R})$ and

$$Mf(x) = \sup_{r>0} (2r)^{-1} \int_{(x-r, x+r)} |f| dm_{id} \text{ for all } x \in \mathbf{R},$$

then

$$m_{id}(\{x \in \mathbf{R} : Mf(x) > \lambda\}) \leq 2\lambda^{-1} \int_{\mathbf{R}} |f| dm_{id} \text{ for all } \lambda > 0.$$

To validate this result of Godfrey Harold Hardy and John Edensor Littlewood, given $f \in LRS_{id}(\mathbf{R})$ let

$$E_\lambda(f) = \{x \in \mathbf{R} : Mf(x) > \lambda\}.$$

Then $E_\lambda(f)$ is open - in fact - if $x \in E_\lambda(f)$, then there is $r > 0$ such that

$$(2r)^{-1} \int_{(x-r, x+r)} |f| dm_{id} > \lambda.$$

Upon choosing $s > r$ such that

$$(2s)^{-1} \int_{(x-r, x+r)} |f| dm_{id} > \lambda,$$

we get that

$$|y - x| < s - r \Rightarrow (x - r, x + r) \subseteq (y - s, y + s)$$

and

$$\lambda < (2s)^{-1} \int_{(x-r, x+r)} |f| dm_{id} \leq (2s)^{-1} \int_{(y-s, y+s)} |f| dm_{id} \leq Mf(y),$$

namely, $y \in E_\lambda(f)$ and so $E_\lambda(f)$ is an open set. Accordingly, any $x \in E_\lambda(f)$ has an open interval $I_x = (x - \rho, x + \rho) \subseteq E_\lambda(f)$ enjoying

$$x \in I_x \text{ and } (m_{id}(I_x))^{-1} \int_{I_x} |f| dm_{id} > \lambda.$$

Consequently, $E_\lambda(f) = \cup_{x \in E_\lambda} I_x$. Since

$$f \in LRS_{id}(\mathbf{R}) \Rightarrow \varepsilon = \lambda^{-1} \int_{\mathbf{R}} |f| dm_{id} > m_{id}(I_x) > 0,$$

an application of Theorem 5.1.3(iv) yields a finite family of mutually disjoint open intervals $\{J_j\}_{j=1}^n$ such that

$$E_\lambda(f) = (E_\lambda(f) \setminus \cup_{j=1}^n J_j) \cup (\cup_{j=1}^n J_j) \text{ and } m_{id}(E_\lambda(f) \setminus \cup_{j=1}^n J_j) < \varepsilon,$$

thereby producing the desired estimate

$$\begin{aligned} m_{id}(E_\lambda(f)) &\leq \sum_{j=1}^n m_{id}(J_j) + \varepsilon \\ &\leq \lambda^{-1} \sum_{j=1}^n \int_{J_j} |f| dm_{id} + \varepsilon \\ &\leq \lambda^{-1} \int_{\cup_{j=1}^n J_j} |f| dm_{id} + \lambda^{-1} \int_{\mathbf{R}} |f| dm_{id} \\ &\leq 2\lambda^{-1} \int_{\mathbf{R}} |f| dm_{id}. \end{aligned}$$

Step 2 - Lebesgue point determination: m_{id} -a.e. $x \in \mathbf{R}$ is Lebesgue's point of $f \in LRS_{id}(\mathbf{R})$.

To do so, for $x \in \mathbf{R}$ let

$$\tilde{M}f(x) = \limsup_{r \rightarrow 0} (2r)^{-1} \int_{(x-r, x+r)} |f - f(x)| dm_{id}.$$

Note that $\tilde{M}f(x) = 0$ implies that x is Lebesgue's point of f . So, for $\lambda \geq 0$ we consider

$$F_\lambda(f) = \{x \in \mathbf{R} : \tilde{M}f(x) > \lambda\}$$

and prove $m_{id}(F_\lambda(f)) = 0$ whose case $\lambda = 0$ implies that m_{id} -a.e. x is Lebesgue's point of f .

Now, for any open interval I with centre x we have

$$(m_{id}(I))^{-1} \int_I |f - f(x)| dm_{id} \leq Mf(x) + |f(x)|,$$

thereby getting

$$2\lambda < Mf(x) + |f(x)| \text{ for all } x \in F_{2\lambda}(f) \text{ and } \lambda > 0.$$

Thus, $\lambda > 0$ and $x \in F_{2\lambda}(f)$ ensure either $Mf(x) > \lambda$ or $|f(x)| > \lambda$. Consequently, an application of *Step 1* derives

$$\begin{aligned} m_{id}(F_{2\lambda}(f)) &\leq m_{id}(\{x \in \mathbf{R} : Mf(x) > \lambda\}) + m_{id}(\{x \in \mathbf{R} : |f(x)| > \lambda\}) \\ &\leq 3\lambda^{-1} \int_{\mathbf{R}} |f| dm_{id}. \end{aligned}$$

In accordance with Theorem 2.4.12 and the Fubini theorem, for any $\varepsilon > 0$ there are $h = f * \phi_n \in C(\mathbf{R})$ and $\eta > 0$ such that

$$\int_{\mathbf{R}} |f - h| dm_{id} \leq \int_{[-\eta, \eta]} |f - f * \phi_n| dm_{id} + 2 \left(\int_{\mathbf{R}} |f| dm_{id} \right) \int_{\mathbf{R} \setminus [-\eta, \eta]} \phi_n dm_{id} < \varepsilon.$$

This, along with Theorem 5.5.1(i), implies

$$\tilde{M}f(x) \leq \tilde{M}(f - h)(x) + \tilde{M}h(x) = \tilde{M}(f - h)(x),$$

and so, if $\lambda > 0$ then

$$\begin{aligned} m_{id}(F_{2\lambda}(f)) &\leq m_{id}(\{x \in \mathbf{R} : \tilde{M}(f - h)(x) > 2\lambda\}) \\ &\leq 3\lambda^{-1} \int_{\mathbf{R}} |f - h| dm_{id} \\ &\leq 3\lambda^{-1} \varepsilon. \end{aligned}$$

Letting $\varepsilon \rightarrow 0$ gives

$$m_{id}(F_{2\lambda}(f)) = 0 \text{ for all } \lambda > 0,$$

whence

$$m_{id}(F_0(f)) = m_{id}(\cup_{0 < 2\lambda \in \mathbf{Q}} F_{2\lambda}(f)) \leq \sum_{0 < 2\lambda \in \mathbf{Q}} m_{id}(F_{2\lambda}(f)) = 0.$$

□

Problems

5.1. For a closed interval $[a, b]$ let $E \subseteq [a, b]$ be m_{id} -measurable and $\varepsilon > 0$. Prove that:

- (i) there is a closed set $C \subseteq E$ such that $0 \leq m_{id}(E) - m_{id}(C) < \varepsilon$;
- (ii) there are an $N \in \mathbf{N}$ and an increasing sequence of closed sets $\{C_j\}_{j=1}^{\infty}$ such that $C_j \subseteq E$ and $0 \leq m_{id}(E) - m_{id}(C_j) < \varepsilon$ when $j \geq N$;
- (iii) $m_{id}(E) = \inf\{m_{id}(O) : \text{open } O \supseteq E\} = \sup\{m_{id}(C) : \text{closed } C \subseteq E\}$. Conversely, if this last equation is valid, then E is m_{id} -measurable.

5.2. Prove that if $E, F \subseteq \mathbf{R}$ then $m_{id,*}$ enjoys the strong sub-additivity

$$m_{id,*}(E \cup F) + m_{id,*}(E \cap F) \leq m_{id,*}(E) + m_{id,*}(F)$$

with equality when E, F are m_{id} -measurable.

5.3. For any subset E of a given compact interval $[a, b]$ let $m_{*,id}(E) = \sup\{m_{id}(C) : \text{closed } C \subseteq E\}$ be the inner measure of E . Prove $m_{*,id}(E) = b - a - m_{id,*}([a, b] \setminus E)$.

5.4. Verify that given $\kappa \in \mathbf{R}$:

- (i) $\limsup_{0 < x-c \rightarrow 0} f(x) \leq \kappa$ if and only if for any $\varepsilon > 0$ there is $\delta > 0$ such that $f(x) < \kappa + \varepsilon$ for all x with $0 < x - c < \delta$;
- (ii) $\liminf_{0 < x-c \rightarrow 0} f(x) \leq \kappa$ if and only if for any $\varepsilon, \delta > 0$ there is x with $0 < x - c < \delta$ such that $f(x) < \kappa + \varepsilon$;
- (iii) if $\overline{D^+}f(c) > R \in \mathbf{R}$ then for any $\varepsilon > 0$ there is $h \in (0, \varepsilon)$ enjoying $R < h^{-1}(f(c+h) - f(c))$;
- (iv) if $\underline{D^+}f(c) < r \in \mathbf{R}$ then for any $\varepsilon > 0$ there is $h \in (0, \varepsilon)$ enjoying $h^{-1}(f(c+h) - f(c)) < r$.

5.5. For a compact interval $[a, b]$ and a function $f : [a, b] \rightarrow \mathbf{R}$ define its positive and negative variation as follows:

$$\begin{cases} \langle V^+ \rangle_{af}^b = \sup \left\{ \sum_{j=1}^n (f(x_j) - f(x_{j-1}))^+ : a = x_0 < x_1 < \cdots < x_n = b \right\}; \\ \langle V^- \rangle_{af}^b = \sup \left\{ \sum_{j=1}^n (f(x_j) - f(x_{j-1}))^- : a = x_0 < x_1 < \cdots < x_n = b \right\}. \end{cases}$$

Prove the following assertions:

$$(i) \max \{ \langle V^+ \rangle_a^f, \langle V^- \rangle_a^b f \} \leq V_a^b f \leq \langle V^+ \rangle_a^b f + \langle V^- \rangle_a^b f;$$

(ii) if $f \in AC[a, b]$, then:

$$\begin{cases} (ii-1) & x \mapsto \langle V^+ \rangle_{[a,x]} f \text{ and } x \mapsto \langle V^- \rangle_{[a,x]} f \text{ are increasing on } [a, b]; \\ (ii-2) & V_{[a,x]} f = \langle V^+ \rangle_{[a,x]} f + \langle V^- \rangle_a^s f \text{ for all } x \in [a, b]; \\ (ii-3) & f(x) - f(a) = \langle V^+ \rangle_{[a,x]} f - \langle V^- \rangle_a^s f \text{ for all } x \in [a, b]. \end{cases}$$

(iii) if $f \in BV[a, b]$, then $\int_{[a,b]} |f'| dm_{id} \leq V_a^b f$ with equality holding for $f \in AC[a, b]$.

(iv) if f' exists everywhere on $[a, b]$ and $f' \in LRS_{id}[a, b]$, then $f \in AC[a, b]$.

5.6. For a compact interval $[a, b]$ let $f \in AC[a, b]$ and $E \subseteq [a, b]$ be m_{id} -measurable. Prove that $f(E)$ is m_{id} -measurable.

5.7. Verify Lebesgue's decomposition theorem on a given compact interval $[a, b]$ - if $f : [a, b] \rightarrow \mathbf{R}$ is increasing, then there are two increasing functions g, h on $[a, b]$ such that $f = g + h$, where $g \in AC[a, b]$ and $h' = 0$ m_{id} -a.e. on $[a, b]$ - moreover - g, h are unique up to a constant.

5.8. Given a compact interval $[a, b]$. Prove that if $f \in LRS_{id}[a, b]$ then

$$\int_{[a,x]} f dm_{id} = 0 \text{ for all } x \in [a, b] \Leftrightarrow f(t) = 0 \text{ for } m_{id} - \text{a.e. } t \in [a, b].$$

5.9. Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be locally m_{id} -integrable and $0 < \alpha \leq 1$. Prove that

$$\sup_{x \neq y \text{ in } \mathbf{R}} \frac{|f(y) - f(x)|}{|y - x|^\alpha} < \infty \Leftrightarrow \sup_{(x,r) \in \mathbf{R} \times (0, \infty)} (2r)^{-1-\alpha} \int_{(x-r, x+r)} |f - f(x)| dm_{id} < \infty.$$

5.10. Verify the following results.

(i) If $f_1 = 1_{[0, \infty)}$, then

$$Mf_1 = \begin{cases} 1 & , x \in (0, \infty) \\ 2^{-1} & , x \in (-\infty, 0] \end{cases}.$$

(ii) If

$$f_2(x) = \begin{cases} (x \ln^2 x)^{-1}, & x \in (0, 2^{-1}) \\ 0 & , \quad x \in \mathbf{R} \setminus (0, 2^{-1}) \end{cases} ,$$

then Mf is not locally m_{id} -integrable.

(iii) If $1 < p < \infty$, then

$$\int_{\mathbf{R}} (Mf)^p dm_{id} \leq p(p-1)^{-1} 2^{p+1} \int_{\mathbf{R}} |f|^p dm_{id}.$$

Chapter 6

Metric Spaces

The objective of this chapter is to establish a basis which is general enough to acquaint ourselves with some limit concepts in such an abstract setting as functional analysis. In this case the word “abstract” implies that the objects we deal with do not have to be numbers or even have arithmetic properties like numbers. The one property that is essential to our idea of limits is the concept of distance between elements. Each limit concept that we have seen thus far is based on the question of whether two numbers are close enough to each other whenever certain conditions hold. So our abstraction is developed in a system whose elements need possess only the property of distance between pairs of elements.

6.1 Metrizable Topology and Connectedness

To begin with, we introduce the definition of a metric space followed by some useful examples, and consider topological properties such as open and closed sets induced by a metric space.

Definition 6.1.1. *A metric space is a system (X, d) that consists of a set X whose elements are called points and a function d whose domain is all of $X \times X$ and whose range is in $[0, \infty)$ such that the following three properties hold:*

- (i) $d(x, y) \geq 0$ for all $x, y \in X$ and $d(x, y) = 0$ if and only if $x = y$;
- (ii) $d(x, y) = d(y, x)$ for every $x, y \in X$;
- (iii) $d(x, z) \leq d(x, y) + d(y, z)$ for every $x, y, z \in X$.

Remark 6.1.2.

(i) Thanks to these three properties of which (ii) and (iii) mean respectively the symmetry and the triangle inequality of the metric space, the function d is usually called the distance function.

(ii) If $d(\cdot, \cdot)$ has all the properties (i)-(ii)-(iii) except that $d(x, y) = 0$ does not necessarily enforce $x = y$, then we call d a pseudo-distance and (X, d) a pseudo-metric space.

Example 6.1.3.

(i) For $n \in \mathbf{N}$ let \mathbf{R}^n be the set of all finite number sequences $x = (x_k)_{k=1}^n$ where the real number x_k is called the k th coordinate of x . Define

$$d_n(x, y) = \left(\sum_{k=1}^n |x_k - y_k|^2 \right)^{2^{-1}} \quad \text{for all } x = (x_k)_{k=1}^n, y = (y_k)_{k=1}^n \in \mathbf{R}^n.$$

Then \mathbf{R}^n with this distance function is called the n -dimensional Euclidean space, or, more briefly, the Euclidean n -space. Our immediate goal, however, is to verify that d_n does indeed provide \mathbf{R}^n with a metric distance via proving the Cauchy-Schwarz inequality for \mathbf{R}^n (see also Theorem 2.2.8 for a continuous form):

$$\sum_{k=1}^n |x_k y_k| \leq \left(\sum_{k=1}^n |x_k|^2 \right)^{2^{-1}} \left(\sum_{k=1}^n |y_k|^2 \right)^{2^{-1}}.$$

To see this, consider the quadratic polynomial

$$(0, \infty) \ni t \mapsto \tau(t) = \alpha t^2 + \beta t + \gamma \quad \text{with} \quad \begin{cases} \alpha = \sum_{k=1}^n |x_k|^2; \\ \beta = 2 \sum_{k=1}^n |x_k| |y_k|; \\ \gamma = \sum_{k=1}^n |y_k|^2. \end{cases}$$

Then

$$\tau(t) = \sum_{k=1}^n (|x_k|t + |y_k|)^2 \geq 0.$$

This infers that $\tau(t)$ cannot have two distinct zeros. Therefore, by the quadratic formula, its discriminant $\beta^2 - 4\alpha\gamma \leq 0$. In other words,

$$\left(\sum_{k=1}^n |x_k| |y_k| \right)^2 - \left(\sum_{k=1}^n |x_k|^2 \right) \left(\sum_{k=1}^n |y_k|^2 \right) \leq 0.$$

Using the Cauchy-Schwarz inequality, we derive that if $x, y, z \in \mathbf{R}^n$, then

$$\begin{aligned} (d_n(x, y) + d_n(y, z))^2 &= d_n(x, y)^2 + 2d_n(x, y)d_n(y, z) + d_n(y, z)^2 \\ &\geq \sum_{k=1}^n ((x_k - y_k) + (y_k - z_k))^2 \\ &= d_n(x, z)^2. \end{aligned}$$

Taking the square root of the first and last members, we further obtain

$$d_n(x, z) \leq d_n(x, y) + d_n(y, z),$$

whence verifying that (\mathbf{R}^n, d_n) is a metric space.

(ii) For $n \in \mathbf{N}$ let $\mathbf{Q}^n = \mathbf{Q} \times \mathbf{Q} \times \cdots \times \mathbf{Q} = \{(x_1, \dots, x_n) \in \mathbf{R}^n : x_1, x_2, \dots, x_n \in \mathbf{Q}\}$. Then (\mathbf{Q}^n, d_n) is also a metric space.

(iii) Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be an increasing function, $E \subseteq \mathbf{R}$ an m_g -measurable set, and $p \in (0, \infty]$. If \mathbf{F} is \mathbf{R} or \mathbf{C} , then the Lebesgue-Radon-Stieltjes space $LRS_g^p(E, \mathbf{F})$ is defined as the collection of all \mathbf{F} -valued functions $f = \Re f + i\Im f$ for which $|f| = \sqrt{(\Re f)^2 + (\Im f)^2}$ and

$$\|f\|_{p, m_g, E} = \begin{cases} \left(\int_E |f(x)|^p dm_g(x) \right)^{\frac{1}{p}}, & p \in (0, \infty) \\ \inf \{c \geq 0 : m_g(\{x \in E : |f(x)| > c\}) = 0\}, & p = \infty \end{cases}$$

is finite. If

$$d_{p, m_g, E}(f_1, f_2) = \begin{cases} \|f_1 - f_2\|_{p, m_g, E}^p, & p \in (0, 1) \\ \|f_1 - f_2\|_{p, m_g, E}, & p \in [1, \infty] \end{cases},$$

then $d_{p, m_g, E}(f_1, f_2) = 0$ if and only if $f_1 = f_2$ m_g -a.e. on E ; and (by Minkowski's inequality)

$$d_{p, m_g, E}(f_1, f_2) \leq d_{p, m_g, E}(f_1, f_3) + d_{p, m_g, E}(f_3, f_2)$$

for $f_1, f_2, f_3 \in LRS_g^p(E)$. Accordingly, $(LRS_g^p(E, \mathbf{F}), d_{p, m_g, E})$ is a pseudo-metric space.

Example 6.1.3 leads to a general fact about metric spaces.

Theorem 6.1.4.

(i) Let X be any nonempty set and define the discrete distance $d(x, y)$, $x, y \in X$, by

$$d(x, y) = \begin{cases} 1, & x \neq y \\ 0, & x = y \end{cases}.$$

Then (X, d) is a metric space.

(ii) If (X, d) is a metric space and Y is a subset of X , then (Y, d) is a metric space for which the domain of d is now restricted to $Y \times Y$.

Proof. (i) It is clear that this d satisfies Definition 6.1.1 (i) and (ii). To verify Definition 6.1.1 (iii), consider any points $x, y, z \in X$. If $d(x, z) = 0$, then Definition 6.1.1 (iii) is trivial. If $d(x, z) > 0$, then $x \neq z$ and $d(x, z) = 1$, and hence y cannot equal both x and z , so at least one of the distances $d(x, y)$ and $d(y, z)$ must equal 1. Therefore, Definition 6.1.1 (iii) follows.

(ii) This truth is an immediate consequence of the definition of metric space and the concept of subset. \square

The topology for a space of points provides a theory of cluster points and convergence for sets of points in the space. In a metric space, this convergence theory is based on the concept of distance between points. For more abstract spaces it is based on a system of open sets, and there may be no distance concept associated with the notion of limit. We here take the former approach and use the metric distance function to define certain open sets, which are generalizations of the open intervals in \mathbf{R} .

In the rest of this section, we always assume that X is a metric space with distance function d .

Definition 6.1.5. If $x \in X$ and $r > 0$, then

$$\begin{cases} B_r(x) = B_r^X(x) = \{y \in X : d(x, y) < r\}; \\ \overline{B}_r(x) = \overline{B}_r^X(x) = \{y \in X : d(x, y) \leq r\}; \\ S_r(x) = S_r^X(x) = \{y \in X : d(x, y) = r\}, \end{cases}$$

are called the open ball; closed ball; sphere of radius r about x respectively. A neighborhood of x is any subset of X that contains some open ball about x .

Remark 6.1.6. Note that $B_0(x) = \emptyset$ and $\overline{B}_0(x) = \{x\}$.

Example 6.1.7. Let $X = \mathbf{R}^n$, $x \in \mathbf{R}^n$ and $r > 0$.

- (i) If $n = 1$, then $B_r(x) = (x - r, x + r)$ and $\overline{B}_r(x) = [x - r, x + r]$
- (ii) If $n = 2$, then $B_r(x)$ and $\overline{B}_r(x)$ are the open and closed disks of radius r centered at x , respectively.
- (iii) If $n = 3$, then $B_r(x)$ and $\overline{B}_r(x)$ are the open and closed (3-dimensional) balls of radius r centered at x , respectively.

Definition 6.1.8. Let $A \subseteq X$.

- (i) $x \in X$ is called an interior point of A , denoted $x \in A^\circ$, provided there exists $r > 0$ such that $B_r(x) \subseteq A$.
- (ii) A is said to be open provided $A \subseteq A^\circ$.
- (iii) $x \in X$ is called a cluster point of A , denoted $x \in A'$, provided for every $r > 0$ one has $B_r(x) \cap (A \setminus \{x\}) \neq \emptyset$.
- (iv) $x \in X$ is called an isolated point of A provided there exists an $r_0 > 0$ such that $B_{r_0}(x) \cap A = \{x\}$.
- (v) A is said to be closed provided $A' \subseteq A$.
- (vi) A is said to be bounded provided $A \subseteq B_r(x)$ for certain $r > 0$ and some $x \in X$.

Remark 6.1.9. It is clear that A is open or closed if and only if $A = A^\circ$ or $A = \overline{A} = A \cup A'$.

Example 6.1.10.

- (i) Any open ball $B_r(x)$ in a given metric space X is a bounded open set.
- (ii) In \mathbf{R}^2 , let $A = B_2(0)$ and $x = (\sqrt{2}, \sqrt{2})$; then $x \in A'$. Since

$$(\sqrt{2} - 2^{-1}\epsilon, \sqrt{2} - 2^{-1}\epsilon) \in B_2(0) \cap B_\epsilon(x)$$

where $B_\epsilon(x)$ is an open disk about x with radius $\epsilon > 0$. However, in \mathbf{R}^2 , if $A = B_1(0) \cup \{x\}$ where $x = (2, 0)$. Then $x \notin A'$, for the open disk $B_{2^{-1}}(x)$ contains no point of A except x itself.

- (iii) Any singleton $\{x\}$ of X is closed.
- (iv) X and \emptyset are not only open, but also closed.

(v) Let $F = \{(1, 0), (2^{-1}, 0), (3^{-1}, 0), \dots, (k^{-1}, 0), \dots\}$ be the subset of \mathbf{R}^2 . Then $(0, 0)$ is a cluster point of F , and since $(0, 0)$ is not in F we infer that F is not closed. Also, the point $(1, 0)$ is not an interior point of F , so F is not open.

The forthcoming theorem is the fundamental relationship between the notions of open set and closed set.

Theorem 6.1.11. *For a subset A of X let $A^c = X \setminus A$ be the complement of a set A in X . Then A is open if and only if A^c is closed.*

Proof. Suppose A is open. To show that A^c is closed, we use the contrapositive method to prove that if $x \notin A^c$ then x is not a cluster point of A^c . Suppose now $x \notin A^c$. Then $x \in A$. Since A is open, there is a ball $B_r(x) \subseteq A$. Thus x is not a cluster point of A^c , as desired.

Conversely, suppose that A^c is closed. If A is not open. Then A contains some point y that is not an interior point of A . Thus every ball $B_r(y)$ contains a point of A^c . Consequently, y is a cluster point of A^c , and since y is in $A \setminus A^c$, we conclude that A^c is not closed, a contradiction to the assumption. \square

Remark 6.1.12. *Perhaps the greatest value of Theorem 6.1.11 is that it gives us an alternative method for proving that a set is open or closed. It is sometimes easier to work with the complement of a set than with the set itself. For instance, we know immediately that $X \setminus \{x\}$ is open because $\{x\}$ is closed.*

Theorem 6.1.13.

- (i) *The intersection of a finite number of open sets is itself an open set.*
- (ii) *The union of any collection of open sets is an open set.*

Proof. (i) Suppose that each of A_1, \dots, A_n is open, and let $x \in A = \bigcap_{k=1}^n A_k$. Then x is in each of A_1, \dots, A_n , so for each $1 \leq k \leq n$ there exists a positive radius r_k such that $B_{r_k}(x) \subseteq A_k$. Define $r = \min\{r_1, \dots, r_n\}$. Then for each k , $B_r(x) \subseteq B_{r_k}(x) \subseteq A_k \subseteq A$. Hence x is in A° , and so A is open.

(ii) Suppose that A_j is an open set for each j in an index set I . Let $x \in \bigcup_{j \in I} A_j$. Then x belongs to some A_j , say, A_{j_0} . Since it is open, there exists an open ball $B_r(x)$ such that $B_r(x) \subseteq A_{j_0} \subseteq \bigcup_{j \in I} A_j$. Hence $\bigcup_{j \in I} A_j$ is open. \square

This theorem leads to a result regarding closed sets.

Corollary 6.1.14.

- (i) *The union of a finite number of closed sets is a closed set.*
(ii) *The intersection of any collection of closed sets is itself a closed set.*

Proof. It suffices to prove (i). Suppose that F_1, \dots, F_n are closed sets. It is easy to verify that $X \setminus \bigcup_{k=1}^n F_k = \bigcap_{k=1}^n (X \setminus F_k)$. By Theorem 6.1.13, each $X \setminus F_k$ is open. Thus, according to Theorem 6.1.13 (i) their intersection is open. Hence $\bigcup_{k=1}^n F_k$ is closed since its complement is open. \square

Remark 6.1.15.

(i) *The union of an infinite collection of closed sets need not be closed. For instance, let F_k be the singleton $(k^{-1}, 0)$ in \mathbf{R}^2 . It is clear that $(0, 0)$ is a cluster point of $\bigcup_{k=1}^{\infty} F_k$, but $(0, 0)$ is not in this union. So the union is not closed. Of course, it is not hard to give an example of an infinite collection of open sets whose intersection is not open. For example, $A_k = B_{k^{-1}}(0)$. Clearly, $\bigcap_{k=1}^{\infty} A_k = \{0\}$ which is not open.*

(ii) *We should take particular notice of the fact that the collections of sets in Theorem 6.1.13 and Corollary 6.1.14 may be arbitrarily large. In particular, the collections may contain so many sets that it is impossible to describe them as an infinite sequence of sets. This is why we use the device of an index set I to describe the collection. Had we written $\{A_n\}_{n=1}^{\infty}$ for the collection of sets, this would have indicated a sequence of sets and therefore limited our discussion to countable collections.*

We finish this section with applying the formula

$$\bar{A} = \{x \in X : B_r(x) \cap A \neq \emptyset \text{ for all } r > 0\}$$

to discussing the connectedness of a metric space – the phenomenon of a set that cannot be naturally separated into two subsets.

The vagueness of the phrase “naturally separated” suggests that it is not easy to give a precise definition of the property we want. Indeed, we define it by first stating when a set does not possess this property.

Definition 6.1.16. *The set $S \subseteq X$ is disconnected provided that there exist nonempty sets S_1 and S_2 such that $S_1 \cup S_2 = S$, $\bar{S}_1 \cap S_2 = \emptyset$ and $\bar{S}_2 \cap S_1 = \emptyset$. If S is not disconnected, then S is said to be connected.*

It is worth pointing out that the above S_1 and S_2 satisfy a property that is stronger than being disjoint. Not only must S_1 and S_2 fail to contain any point of the other set, they must also fail to contain any cluster point of the other set. Note that this extra stipulation in Definition 6.1.16 is significant because any set containing two or more points can be written as the union of two nonempty disjoint subsets – for $x \in S$, let $S_1 = \{x\}$ and $S_2 = S \setminus \{x\}$. But as we see in the next theorem, when we are dealing with open sets, disjointness is sufficient to imply that their union is disconnected.

Theorem 6.1.17. *Let $S \subseteq X$. Then S is disconnected if and only if there are disjoint open sets A and B such that $S \subseteq A \cup B$, $A \cap S \neq \emptyset$, and $B \cap S \neq \emptyset$.*

Proof. Suppose that A and B exist as in the statement of Theorem 6.1.17 and let $S_1 = A \cap S$ and $S_2 = B \cap S$. Clearly, S_1 and S_2 are nonempty, and $S_1 \cup S_2 = S$. Consider an arbitrary point $x \in S_1$; then x is in the open set A , so for some $r > 0$, $B_r(x) \subseteq A$. Thus $B_r(x) \cap B = \emptyset$ because $A \cap B = \emptyset$, so $B_r(x) \cap S_2 = \emptyset$. Thus $x \notin \overline{S_2}$; hence $S_1 \cap \overline{S_2} = \emptyset$. Similarly, $\overline{S_1} \cap S_2 = \emptyset$, and we conclude that S is disconnected.

Now assume that S is disconnected, say, $S = S_1 \cup S_2$ and $S_1 \cap \overline{S_2} = \overline{S_1} \cap S_2 = \emptyset$. Then no point of S_1 is in S'_2 . For $x \in S_1$, choose $B_r(x)$ such that $B_r(x) \cap S_2 = \emptyset$. Thus $A = \bigcup_{x \in S_1} B_{2^{-1}r}(x)$ is an open set that contains no point of S_2 . Similarly, we can take a ball $B_s(y)$ for each $y \in S_2$ that satisfies $B_s(y) \cap S_1 = \emptyset$ and define $B = \bigcup_{y \in S_2} B_{2^{-1}s}(y)$, which is open. It remains only to show that $A \cap B = \emptyset$. Suppose not, and let $p \in A \cap B$. Then for some $x \in S_1$ and $y \in S_2$, $p \in B_{2^{-1}r}(x) \cap B_{2^{-1}s}(y)$. But this implies that

$$d(x, y) \leq d(x, p) + d(p, y) < 2^{-1}(r + s) \leq \max\{r, s\}.$$

Therefore either $y \in B_r(x)$ or $x \in B_s(y)$, which contradicts the choice of either r or s , respectively. \square

Corollary 6.1.18. *Let $S \subseteq X$. If S is connected, then \overline{S} is connected.*

Proof. Assume that \overline{S} is disconnected, and let A and B be disjoint open sets in Theorem 6.1.17: $\overline{S} \subseteq A \cup B$, $A \cap \overline{S} \neq \emptyset$, and $B \cap \overline{S} \neq \emptyset$. Since S is obviously contained in $A \cup B$, we need to prove only that $A \cap S \neq \emptyset$ and $B \cap S \neq \emptyset$. Let $x \in \overline{S} \cap A$. Since A is open, there is a ball about x such that $B_r(x) \subseteq A$. But $x \in \overline{S}$ too, so $B_r(x)$ must contain some point y of S . Therefore $y \in B_r(x) \cap S \subseteq A \cap S$, so $A \cap S \neq \emptyset$. Similarly, $B \cap S \neq \emptyset$, so by Theorem 6.1.17, S is disconnected. This contradicts the assumption that S is connected. \square

6.2 Completeness

To discuss the completeness of a metric space, we start with the definition of the convergence of a point sequence in the metric space.

A point sequence is a function from \mathbf{N} into a metric space (X, d) ; this is, it is a sequence whose terms are points in X . Since our principal examples of metric spaces are the Euclidean spaces and we use subscripts to designate the coordinates of points in \mathbf{R}^n , we use superscripts to index the terms of a point sequence — $\{x^{(k)}\}_{k=1}^{\infty}$ represents a point sequence in a metric space.

Definition 6.2.1. *Given a metric space (X, d) . The point sequence $\{x^{(k)}\}_{k=1}^{\infty}$ in X is said to converge to the point $x \in X$, denoted $\lim_{k \rightarrow \infty} x^{(k)} = x$, provided for every $\varepsilon > 0$ there is an $N \in \mathbf{N}$ such that $d(x^{(k)}, x) < \varepsilon$ whenever $k > N$.*

Our first task is to establish a connection between convergent point sequences and cluster points.

Theorem 6.2.2. *Given a metric space (X, d) and $A \subseteq X$. Then $x \in A'$ if and only if there is a non-repeating point sequence in A that converges to x .*

Proof. Suppose $x \in A'$ and $r_1 = 1$. Then there exists a point $x^{(1)}$ in $B_{r_1}(x) \cap (A \setminus \{x\})$. Let $r_2 = \min\{2^{-1}, d(x^{(1)}, x)\}$, and choose a point $x^{(2)}$ in $(A \setminus \{x\}) \cap B_{r_2}(x)$. After $x^{(1)}, \dots, x^{(k)}$ have been defined, let $r_{k+1} = \min\{(k+1)^{-1}, d(x^{(k)}, x)\}$ and choose $x^{(k+1)}$ as a point of $A \setminus \{x\}$ in the open ball $B_{r_{k+1}}(x)$. Since $r_k \leq k^{-1}$, it is clear that $\lim_{k \rightarrow \infty} x^{(k)} = x$; and since $r_k \leq d(x^{(k-1)}, x)$, it follows that no two $x^{(k)}$'s are the same.

To prove the converse, we simply observe that if A contains such a point sequence that converges to x , then it is obvious from the definition that every open ball about x contains infinitely many points of A . \square

Example 6.2.3. *In \mathbf{R}^n there is a very strong connection between convergent point sequences and convergent number sequences. More precisely, $x^{(k)} \rightarrow x$ in \mathbf{R}^n if and only if $\lim_{k \rightarrow \infty} x_j^{(k)} = x_j$ for each $j = 1, \dots, n$. This is because $x^{(k)} \in B_\varepsilon(x)$ if and only if*

$$d_n(x^{(k)}, x) = \left(\sum_{j=1}^n |x_j^{(k)} - x_j|^2 \right)^{2^{-1}} < \varepsilon.$$

Definition 6.2.4. Suppose (X, d) is a metric space. A point sequence $\{x^{(k)}\}_{k=1}^{\infty}$ in X is a Cauchy sequence provided for every $\varepsilon > 0$ there is an $N > 0$ such that $d(x^{(k)}, x^{(m)}) < \varepsilon$ whenever $k > m > N$.

This definition is obviously a generalization of the concept of Cauchy number sequences treated in Chapter 1.

Example 6.2.5. A point sequence in \mathbf{R}^n is a Cauchy sequence if and only if it converges to a point in \mathbf{R}^n . To verify this, suppose that $\{x^{(k)}\}_{k=1}^{\infty}$ is a Cauchy sequence. Since

$$d_n(x^{(k)}, x^{(m)}) \geq |x_j^{(k)} - x_j^{(m)}| \text{ for each } j = 1, \dots, n,$$

it follows that $\{x_j^{(k)}\}_{k=1}^{\infty}$ forms a Cauchy number sequence. Thus such a sequence converges to a point in \mathbf{R} , say, $\lim_{k \rightarrow \infty} x_j^{(k)} = x_j$. Now Example 6.2.3 ensures that $\{x^{(k)}\}_{k=1}^{\infty}$ converges to $x = (x_1, \dots, x_n)$ in \mathbf{R}^n . The converse assertion is true in any metric space, and therefore we give it as the next theorem.

Theorem 6.2.6. Suppose (X, d) is a metric space.

- (i) If $\{x^{(k)}\}_{k=1}^{\infty}$ in X is convergent, then $\{x^{(k)}\}_{k=1}^{\infty}$ is a Cauchy sequence.
- (ii) If $\{x^{(k)}\}_{k=1}^{\infty}$ in X is a Cauchy sequence, then it is bounded.
- (iii) If $x \in X$ and $Y = X \setminus \{x\}$, then any sequence $\{x^{(k)}\}_{k=1}^{\infty}$ in Y with $\lim_{k \rightarrow \infty} d(x^{(k)}, x) = 0$ is a Cauchy but not a convergent sequence in Y with respect to d .

Proof. (i) Assume that $\{x^{(k)}\}_{k=1}^{\infty}$ converges to x , and suppose $\varepsilon > 0$. Choose $N > 0$ so that $d(x^{(k)}, x) < 2^{-1}\varepsilon$ as $k > N$. This implies that

$$d(x^{(k)}, x^{(m)}) \leq d(x^{(k)}, x) + d(x^{(m)}, x) < \varepsilon,$$

whenever $k > m > N$. Therefore, $\{x^{(k)}\}_{k=1}^{\infty}$ is a Cauchy sequence in X .

(ii) Let $\{x^{(k)}\}_{k=1}^{\infty}$ be a Cauchy sequence. By definition, we find an $N > 0$ such that $d(x^{(k)}, x^{(m)}) < 1$ whenever $k > m > N$. In particular, if $m = N + 1$ this property becomes $d(x^{(k)}, x^{(N+1)}) < 1$ whenever $k > N$. Therefore all the points $x^{(k)}$ for $k > N$ lie in the open ball $B_1(x^{(N+1)})$. Now we simply enlarge the radius of the ball until it also includes the first N points of the sequence. Define

$$r = 1 + \max\{1, d(x^{(1)}, x^{(N+1)}), \dots, d(x^{(N)}, x^{(N+1)})\}.$$

Thus r is at least one unit more than the distance between $x^{(N+1)}$ and $x^{(k)}$ for every $k = 1, 2, \dots$; so $B_r(x^{(N+1)})$ contains every point $x^{(k)}$ of the sequence.

(iii) Suppose that $x \in X$ is the limit of $\{x^{(k)}\}_{k=1}^\infty$ under d . Then $\{x^{(k)}\}_{k=1}^\infty$ is a sequence in both X and Y , and it converges in X ; so by (i) it is a Cauchy sequence in X . The distance function d on $Y \times Y$ is the same one as on $X \times X$, so $\{x^{(k)}\}_{k=1}^\infty$ is also a Cauchy sequence in Y . But in Y there is no point to which $\{x^{(k)}\}_{k=1}^\infty$ can converge, because x has been removed and the limit of a sequence is unique. \square

Accordingly, we see from Theorem 6.2.6 that Example 6.2.5 cannot be generalized to all metric spaces. This idea is the general principle behind the following example.

Example 6.2.7.

(i) Let $X = B_1(0)$ be the open unit ball in \mathbf{R}^2 with the usual Euclidean distance. Consider a sequence in $B_1(0)$ that converges to a point on its boundary, say, $x^{(k)} = (1 - k^{-1}, 0)$, so that it has cluster point $(1, 0)$. Then this sequence is a Cauchy sequence in X , but it has no limit in X , so it does not converge in X .

(ii) Again for $n \in \mathbf{N}$ let \mathbf{Q}^n be the subset of \mathbf{R}^n consisting of all points for which all coordinates are rational numbers. The point $x = (\sqrt{2}, 0, \dots, 0)$ is in $\mathbf{R}^n \setminus \mathbf{Q}^n$, but if $\{r_k\}_{k=1}^\infty$ is a sequence of rational numbers such that $\lim_{k \rightarrow \infty} r_k = \sqrt{2}$, and $x^{(k)} = (r_k, 0, \dots, 0)$, then $\lim_{k \rightarrow \infty} x^{(k)} = x$. Therefore $\{x^{(k)}\}_{k=1}^\infty$ is a Cauchy sequence but does not converge in \mathbf{Q}^n .

Naturally, we can give the definition of the completeness of a metric space.

Definition 6.2.8. Suppose (X, d) is a metric space. $S \subseteq X$ is said to be complete with respect to d provided every Cauchy sequence in S converges to a point in S with respect to d . In particular, if X is complete with respect to d , then we say that (X, d) is a complete metric space.

Theorem 6.2.9. Let (X, d) be a complete metric space. Then $S \subseteq X$ is complete with respect to d if and only if S is a closed subset of X .

Proof. In fact, suppose that S is closed and $\{x_k\}_{k=1}^\infty$ is a Cauchy sequence in S . Since X is complete, there is an $x \in X$ such that $x_k \rightarrow x$. But because S is closed, one gets $x \in S$. Conversely, if S is complete, then for any sequence $\{x_k\}_{k=1}^\infty$ converging to $x \in X$, one has that this sequence is a Cauchy sequence in S and, therefore, converges to a point $y \in S$. So $x = y \in S$ and S is closed. \square

In the terminology of Definition 6.2.8, \mathbf{R}^n is a complete metric space. The next result gives four alternative properties that are equivalent to completeness in \mathbf{R}^n . Each is an important result in its own right, and collectively they are the most important tools of analysis. They are immediately recognizable as multidimensional extensions of the four theorems of \mathbf{R}^1 -case.

Theorem 6.2.10. *For $n \in \mathbf{N}$ let $X = \mathbf{R}^n$ be equipped with the distance $d_n(\cdot, \cdot)$. Then the following statements are equivalent:*

- (i) *X is complete.*
- (ii) *The Heine-Borel property: If F is a closed and bounded subset of X and $\{A_\mu\}_{\mu \in I}$ is a collection of open sets whose union contains F , then there is a finite subcollection $\{A_{\mu_j}\}_{j=1}^m$ such that $F \subseteq \cup_{j=1}^m A_{\mu_j}$.*
- (iii) *The Bolzano-Weierstrass property: If S is an infinite bounded subset of X , then $S' \neq \emptyset$.*
- (iv) *The bounded sequence property: If $\{x^{(k)}\}_{k=1}^\infty$ is a bounded sequence in X , then it has a convergent subsequence.*
- (v) *The nested set property: If $\{F_k\}_{k=1}^\infty$ is a sequence of closed and bounded nonempty sets such that $F_{k+1} \subseteq F_k \subseteq X$ for each $k \in \mathbf{N}$, then $\cap_{k=1}^\infty F_k \neq \emptyset$.*

Proof. Since statements (ii)-(v) refer to bounded sets, it will be more convenient to use another form of the boundedness; that is, a set S is bounded in X if and only if it is contained in an n -cube Q with side-length $2r > 0$, i.e.,

$$S \subseteq Q = \{x \in \mathbf{R}^n : |x_j| \leq r \text{ for } j = 1, \dots, n\}.$$

(i) \Rightarrow (ii) Let F be a closed and bounded set and suppose that we are given that $\{A_\mu\}_{\mu \in I}$ is an open cover of F as in (ii). Since F is bounded, there is an n -cube Q_1 containing F . Assume that F cannot be covered by any finite number of the A_μ 's. Subdivide Q_1 into 2^n n -cubes by halving each of the coordinate intervals of Q_1 ; that is, the j th coordinate of a point x in one of the n -cubes is restricted to either $[0, r]$ or $[-r, 0]$ instead of $[-r, r]$. If it were possible to cover those parts of F that lie in each one of the 2^n n -cubes with finitely many A_μ 's, then the union of all of these A_μ 's would still be only a finite number and would cover F . Therefore one of these 2^n n -cubes must contain a subset of F that cannot be covered by any finite number of the A_μ 's. Call this n -cube Q_2 . Now subdivide Q_2 into 2^n n -cubes and repeat the process. The result is a sequence of n -cubes such that $Q_{k+1} \subseteq Q_k$, and the j th coordinate of each

point in Q_k is restricted on an interval whose length is $r2^{2-k}$. Choose a point sequence $\{x^{(k)}\}_{k=1}^\infty$ such that $x^{(k)}$ is in $F \cap Q_k$. Then $\{x_j^{(k)}\}_{k=1}^\infty$ — the sequence of j th coordinates forms a Cauchy number sequence in \mathbf{R} thanks to

$$|x_j^{(k)} - x_j^{(m)}| \leq r(2^{2-k} + \dots + 2^{2-m}) < r2^{3-m} \rightarrow 0 \text{ as } k > m \rightarrow \infty.$$

Consequently, $\{x^{(k)}\}_{k=1}^\infty$ is a Cauchy sequence in \mathbf{R}^n . Assuming (i), we conclude that there is a point $x \in \mathbf{R}^n$ such that $\lim_{k \rightarrow \infty} x^{(k)} = x$. Since each $x^{(k)}$ is in F and F is closed, we have $x \in F$. Therefore x is in one of the A_μ 's, say, $x \in A_{\mu_x}$. But A_{μ_x} is open, so for some $s > 0$, $x \in B_s(x) \subseteq A_{\mu_x}$, which means that $B_s(x)$ contains all but a finite number of the n -cubes Q_k 's. But this derives that the points of F in all these Q_k 's are contained in A_{μ_x} , and this contradicts the choice of Q_k .

(ii) \Rightarrow (iii) Suppose that S is a bounded set which has no cluster point. We show that S must be finite, which establishes (iii). Let \bar{B} be a closed and bounded ball containing S . Since no point of \bar{B} is a cluster point of S , we can choose an open ball $B_{r_x}(x)$ for each $x \in \bar{B}$ such that $B_{r_x}(x)$ contains no point of S except possibly x itself. Now $\bar{B} \subseteq \bigcup_{x \in \bar{B}} B_{r_x}(x)$, so $\{B_{r_x}(x) : x \in \bar{B}\}$ is a collection of open sets whose union contains \bar{B} . Therefore from (ii) we conclude that there is a finite subcollection of these open balls that covers \bar{B} . But this subcollection must also cover S due to $S \subseteq \bar{B}$, and since each open ball contains at most one point of S , we conclude that S is finite.

(iii) \Rightarrow (iv) Assume that $\{x^{(k)}\}_{k=1}^\infty$ is a bounded point sequence, and let S be the range of this sequence; that is, $S = \{p \in \mathbf{R}^n : p = x^{(k)} \text{ for some } k\}$. If S is a finite set, then at least one of its points must appear infinitely many times as a term of the sequence, in which case $\{x^{(k)}\}_{k=1}^\infty$ has a constant subsequence. If S has infinitely many points, then (iii) implies that S has a cluster point, and in proving Theorem 6.2.2 we have seen how to construct a sequence in S that converges to that cluster point.

(iv) \Rightarrow (v) Let $\{F_k\}_{k=1}^\infty$ be a nested sequence of closed and bounded sets as in the hypothesis of (v). For each k , choose a point $x^{(k)}$ in F_k . Every such point is in F_1 , so $\{x^{(k)}\}_{k=1}^\infty$ is bounded sequence. By (iv), $\{x^{(k)}\}_{k=1}^\infty$ has a convergent subsequence whose limit we here call x . We see that the nested property of the sets guarantees that each F_k contains all but possibly a finite number of the points $x^{(k)}$. Therefore for each k there is a sequence of points in F_k that converges to x . Since they are closed sets, x must belong to each F_k . Hence $\bigcap_{k=1}^\infty F_k$ contains x , so the intersection is nonempty.

(v) \Rightarrow (i) Let $\{x^{(k)}\}_{k=1}^\infty$ be a Cauchy sequence, and for each m set F_m be the closure of the set of points $\{x^{(k)} : k \geq m\}$. Thus $\{F_m\}_{m=1}^\infty$ is a nested sequence

of closed sets. By Theorem 6.2.6 (ii) we see that $\{x^{(k)}\}_{k=1}^{\infty}$ is bounded, and therefore F_1 is bounded. By (v) there is a point x in all sets F_m 's. Let $B_{\epsilon}(x)$ be any open ball about x , and choose a number N such that $k > m > N$ implies that $d_n(x^{(k)}, x^{(m)}) < 2^{-1}\epsilon$. Since x is in F_N , which is the closure of the subsequence $\{x^{(k)}\}_{k=N}^{\infty}$, there is some point $x^{(m)}$, where $m > N$, such that $x^{(m)}$ is in $B_{2^{-1}\epsilon}(x)$. Now if $k > N$, then

$$d_n(x^{(k)}, x) \leq d_n(x^{(k)}, x^{(m)}) + d_n(x^{(m)}, x) < \epsilon,$$

and hence $\lim_{k \rightarrow \infty} x^{(k)} = x$. □

Remark 6.2.11.

(i) *In examining the proof of Theorem 6.2.10, we find that the Euclidean distance formula was used only in proving (i) \Rightarrow (ii). Therefore we conclude that the other four implications hold in any metric space.*

(ii) *For $x \in \mathbf{R}^n$ let $X = \mathbf{R}^n \setminus \{x\}$ and the distance between points in X be the same Euclidean distance as in \mathbf{R}^n . Now if $\{x^{(k)}\}_{k=1}^{\infty}$ is a sequence that converges to x in \mathbf{R}^n and $x^{(k)} \neq x$, then it is still a Cauchy sequence in X . But because there is no point in X to which it converges, $\{x^{(k)}\}$ is not convergent in X . Thus X is not complete. Since completeness is implied by the other four statements, it follows that all five are false.*

(iii) *Of course, the fact that all five statements in Theorem 6.2.10 are true in $X = \mathbf{R}^n$ is quite dependent upon the distance formula. The use of the distance formula in proving (i) \Rightarrow (ii) is very subtle. It is inherent in the notions of boundedness and n -cubes that were used.*

Example 6.2.12. *Let X be an infinite set and define d by*

$$d(x, y) = \begin{cases} 1, & x \neq y \\ 0, & x = y \end{cases}.$$

Since each point in X can be written as $\{x\} = B_{2^{-1}}(x)$, we see that every singleton set is an open set in X . Also $X = B_2(x)$ for any $x \in X$, so X is bounded. Therefore X is a bounded and closed set, and $\cup_{x \in X} \{x\}$ is an open cover of X that obviously cannot be reduced to a finite number of open sets. We now assert that X is complete, because it is not hard to show that if $\{x^{(k)}\}_{k=1}^{\infty}$ is a Cauchy sequence, then there is an N such that $x^{(k)} = x^{(N)}$ whenever $k \geq N$. Therefore $\lim_{k \rightarrow \infty} x^{(k)} = x^{(N)}$, so X is complete even though the Heine-Borel property does not hold.

6.3 Compactness, Density and Separability

As for an expansion of the Heine-Borel property, we discuss the compactness, density and countability of sets in a given metric space.

The first important notion induced by the Heine-Borel property is that of compact sets. To see this picture clearly, we first consider two examples that enable us to describe some important and attractive results in the theory of functions.

Example 6.3.1. *A set $K \subseteq \mathbf{R}^n$ is said to be compact provided every sequence in K contains a subsequence that converges to a point in K . This notion can be characterized by the equivalence of the following three statements:*

- (i) K is compact;
- (ii) K is closed and bounded;
- (iii) Every open cover of K has a finite subcover; that is, if $K \subseteq \bigcup_{\alpha \in I} G_\alpha$ where each G_α is an open subset of \mathbf{R}^n , then there is a finite subset J of the index set I such that $K \subseteq \bigcup_{\alpha \in J} G_\alpha$.

In fact, suppose that (i) holds. Then every convergent sequence in K converges to a point in K . Using the fact that a subset E of \mathbf{R}^n is closed if and only if E equals its closure, we get that K is closed. If K is not bounded, then for each $k \in \mathbf{N}$, there is an $x_k \in K$ such that $d_n(x_k, 0) > k$. Consequently, no subsequence of $(x_k)_{k=1}^\infty$ can converge and so K is not compact, a contradiction. Thus, (ii) holds.

If (ii) holds, then an application of Theorem 6.2.10 (i) yields (iii).

Finally, suppose that (iii) holds. To verify (i), it suffices to prove that K is bounded and closed since Theorem 6.2.10 (iv) implies that every sequence in such K has a subsequence which converges to a point in K . To see that K is bounded, set $G_x = B_1(x)$. Then $\{G_x : x \in K\}$ is an open cover of K and so has a finite subcover, say, G_{x_1}, \dots, G_{x_k} . Let $M = \max\{d_n(x_j, 0) : j = 1, \dots, k\}$. If $x \in K$, then $x \in G_{x_j}$ for some j and $d(x, 0) \leq d_n(x_j, 0) + 1 \leq M + 1$. Thus, K is bounded. To show that K is closed, fix a point $x \in \mathbf{R}^n \setminus K$. For any $y \in K$, there are disjoint open balls U_y and V_y centered at y and x , respectively. Note that these two balls will generally depend on x . The family $\{U_y : y \in K\}$ is an open cover of K and so has a finite subcover, say, U_{y_1}, \dots, U_{y_k} . Then $K \subseteq \bigcup_{j=1}^k U_{y_j}$, which is disjoint from the open ball $B = \bigcap_{j=1}^k V_{y_j}$ centered at x . This ball B must be contained entirely in $\mathbf{R}^n \setminus K$: Otherwise, $B \cap K \neq \emptyset$ and hence $B \cap (\bigcup_{j=1}^k U_{y_j}) \supseteq B \cap K \neq \emptyset$, a contradiction. It follows that $\mathbf{R}^n \setminus K$ is open and so K is closed.

Example 6.3.2. The family $\mathcal{F} = \{(k^{-1}, 2) : k \in \mathbf{N}\}$ is an open cover of $(0, 1]$; the family $\mathcal{G} = \{(2^{-k}, 2) : k \in \mathbf{N}\}$ is a subfamily of \mathcal{F} . Each point of $(0, 1]$ belongs to $(2^{-k}, 2)$ for k sufficiently large, and so \mathcal{G} is a subcover for $(0, 1]$. If \mathcal{H} is a finite subfamily of \mathcal{F} , then there is a largest value of k , say k_1 such that $(2^{-k_1}, 2) \in \mathcal{H}$. Since $2^{-k_1-1} \in (0, 1]$ and does not belong to any element of \mathcal{H} , \mathcal{H} is not a subcover for $(0, 1]$. A close look at this example reveals that the problem lies in the fact that $(0, 1]$ is not closed. Other examples, which the reader should construct, show that similar problems can arise with unbounded sets.

Remark 6.3.3. There is an important result generalizing the nested set property – if $\mathcal{F} = \{K_j\}_{j \in I}$ is a class of compact subsets of \mathbf{R}^n and the intersection of any finite subclass of \mathcal{F} is nonempty then $\bigcap_{j \in I} K_j \neq \emptyset$. To see this, let $K \in \mathcal{F}$ be such that $K \cap (\bigcap_{j \in I} K_j) = \emptyset$. Then $K \subseteq \bigcup_{j \in I} (\mathbf{R}^n \setminus K_j)$. Since K is compact, there are finitely many indices j_1, \dots, j_k such that

$$K \subseteq \bigcup_{l=1}^k (\mathbf{R}^n \setminus K_{j_l}) = \mathbf{R}^n \setminus \bigcap_{l=1}^k K_{j_l},$$

so $K \cap (\bigcap_{l=1}^k K_{j_l}) = \emptyset$, a contradiction. Accordingly, $K \cap (\bigcap_{j \in I} K_j) \neq \emptyset$ and $\bigcap_{j \in I} K_j \neq \emptyset$.

An advantage of the last characterization of compactness in Example 6.3.1 is that it remains meaningful when the discussion is extended to the topological spaces much more general than \mathbf{R}^n , and thus serves as the basis for a definition of the compactness in these more general spaces.

Definition 6.3.4. Suppose that (X, d) is a metric space. We say that $S \subseteq X$ is compact provided that whenever S is contained in the union of a collection of open subsets of X , S is contained in the union of a finite number of these open subsets. In particular, if $S = X$ then (X, d) is called a compact metric space.

It is very natural to give an alternative of Definition 6.3.4 in terms of cluster point.

Theorem 6.3.5. Let (X, d) be a metric space and $S \subseteq X$. Then the following statements are equivalent:

- (i) S is compact;
- (ii) Any infinite subset of S has at least one cluster point in S ;

(iii) Any sequence of points in S has a subsequence which is convergent to a point in S .

Proof. (i) \Rightarrow (ii) Suppose that (i) is true. If (ii) is false, then there is an infinite subset S of X that has no cluster point in S . Now for each $x \in S$, we can find an open ball having x as center and containing only a finite number of points of S . Note that S is compact and covered by all such open balls centered at x . So it is subset of the union of a finite number of such open balls of which each contains only a finite number of points of S . This implies that S is finite, a contradiction.

(ii) \Rightarrow (iii) This implication is trivial.

(iii) \Rightarrow (i) Suppose that (iii) is valid, but (i) is not true. Given $\varepsilon > 0$ and $x_1 \in S$. Choose $x_2 \in S$ such that $d(x_2, x_1) \geq \varepsilon$. For $k = 1, 2$ choose $x_3 \in S$ such that $d(x_3, x_k) \geq \varepsilon$. Continuing the previous selection, we can find $x_{n+1} \in S$ such that $d(x_{n+1}, x_k) \geq \varepsilon$ for $k = 1, 2, \dots, n$ and $n \in \mathbf{N}$. This process can be done because S is covered by all open balls centered at points in S of radius ε but this cover has no a finite subcover thanks to the assumption that S is not compact. According to (iii), $\{x_n\}_{n=1}^\infty$ must possess a subsequence $\{x_{n_j}\}_{j=1}^\infty$ which is convergent to a point in S , so $d(x_{n_j}, x_{n_k}) < \varepsilon$ whenever n_j and n_k are big enough. This contradicts $d(x_{n_j}, x_{n_k}) \geq \varepsilon$. \square

Here is an interesting consequence of Theorem 6.3.5.

Corollary 6.3.6. *Let (X, d) be a complete metric space. Then a closed subset S of X is compact if and only if S is totally bounded; that is, for any $\varepsilon > 0$ there exist finitely many open balls $\{B_\varepsilon(x_j)\}_{j=1}^n$ in X such that $S \subseteq \bigcup_{j=1}^n B_\varepsilon(x_j)$.*

Proof. Suppose $S \subseteq X$ is closed. If S is compact and $\varepsilon > 0$, then the trivial inclusion $S \subseteq \bigcup_{x \in S} B_\varepsilon(x)$ induces a finite cover $S \subseteq \bigcup_{j=1}^n B_\varepsilon(x_j)$, as desired.

Conversely, suppose that S is totally bounded. Let S be covered by a collection of open sets $\{O_\alpha\}_{\alpha \in I}$. To reach a contradiction, assume S cannot be covered by any finite sub-collection of $\{O_\alpha\}_{\alpha \in I}$. Since S is closed, from the total boundedness of S it follows that for $\varepsilon = 1$ there is finitely many open balls $\{B_1(x_j)\}_{j=1}^{n_1}$ such that $S = \bigcup_{j=1}^{n_1} S \cap \overline{B}_1(x_j)$. According to the assumption, one of closed sets in $\{S \cap \overline{B}_1(x_j)\}_{j=1}^{n_1}$, say, $S_1 = S \cap \overline{B}_1(x_1)$, cannot be covered by finitely many elements of $\{O_\alpha\}_{\alpha \in I}$. Do the same with $\varepsilon = 2^{-1}$ and S_1 in place of $\varepsilon = 1$ and S respectively, and continue this process. The result is a sequence of closed sets $(S_k)_{k=1}^\infty$ such that:

(a) $S \supseteq S_1 \supseteq S_2 \supseteq \dots$;

(b) $\sup_{x,y \in S_k} d(x,y) \leq k^{-1}$;

(c) each S_k cannot be covered by finitely many elements of $\{O_\alpha\}_{\alpha \in I}$.

Choose $x_k \in S_k$. By (a), (b) and the fact that X is complete and each S_k is closed, $\{x_k\}_{k=1}^\infty$ is a Cauchy sequence which is convergent to an element $x \in \bigcap_{k=1}^\infty S_k$. This implies $x \in S_k \subseteq O_\alpha$ for some $\alpha \in I$ and for sufficiently large $k \in \mathbf{N}$, contradicting (c). Therefore, S must be covered by finitely many elements of $\{O_\alpha\}_{\alpha \in I}$, i.e., S is compact. \square

One example of a compact metric space will give us many more, by means of the following result.

Theorem 6.3.7. *Let (X, d) be a compact metric space. Then X is complete with respect to d . Furthermore:*

(i) *Any closed subset S of X is compact;*

(ii) *For any sequence of nonempty closed subsets $(S_j)_{j=1}^\infty$ of X , with $S_1 \supseteq S_2 \supseteq S_3 \supseteq \cdots$, there is at least one point in $\bigcap_{j \in \mathbf{N}} S_j$.*

Proof. From Theorem 6.3.5 (iii) it follows that any Cauchy sequence has a convergent subsequence and therefore is itself convergent. Thus, (X, d) is complete.

(i) Let S be closed. Then S^c is open. If $S \subseteq \bigcup_{\alpha \in I} U_\alpha$ where $U_\alpha \subseteq X$ is open for each $\alpha \in I$. Then $X \subseteq (\bigcup_{\alpha \in I} U_\alpha) \cup S^c$. Since X is compact, we can find a finite subset $J \subseteq I$ such that $X \subseteq (\bigcup_{\alpha \in J} U_\alpha) \cup S^c$. Hence $S \subseteq \bigcup_{\alpha \in J} U_\alpha$. This shows that S is compact.

(ii) If $\bigcap_{j \in \mathbf{N}} S_j = \emptyset$, then $X = \emptyset^c = \bigcup_{j \in \mathbf{N}} S_j^c$. Since X is compact, there is a finite number of the open subsets $\{S_{i_j}^c\}_{j=1}^J$ such that $X = \bigcup_{j=1}^J S_{i_j}^c$. Note that since $S_1^c \subseteq S_2^c \subseteq S_3^c \subseteq \cdots$, we must have $X = S_k^c$ for some k , which produces the contradiction $S_k = X^c = \emptyset$. \square

Remark 6.3.8. *Here, we point out that Theorem 6.3.7 (ii) does not hold if the word “compact” is replaced by “complete” in the hypothesis – for example, $X = \mathbf{R}$ and $S_k = \{x \in \mathbf{R} : x \geq k\}$ for $k \in \mathbf{N}$.*

In addition to this counterexample, we have the following interesting consequence.

Corollary 6.3.9. *Suppose that (X, d) is a metric space. If $S \subseteq X$ is compact, then it is bounded and closed, but not conversely.*

Proof. Since X is the union of its open balls, if $S \subseteq X$ is compact then S is covered by a finite number of those balls and hence S is bounded. To establish the closedness of S , we note that (S, d) is a compact metric space and so it is complete, by Theorem 6.3.7. Accordingly, S is closed.

For the converse, consider the distance d defined by $d(x, y) = 1$ or $d(x, y) = 0$ if $x \neq y$ or $x = y$. Clearly, any infinite set X is bounded and closed under d , but X is not compact since $X = \bigcup_{x \in X} B_{2^{-1}}(x)$ has no finite cover. \square

Next we deal with the density and the separability of metric spaces.

Definition 6.3.10. Suppose that (X, d) is a metric space.

(i) $D \subseteq X$ is said to be dense in X provided that any open ball in X contains a point of D .

(ii) X is called separable provided that there is a countable set $S = \{x_1, x_2, \dots\} \subseteq X$ that is dense in X .

To give an example, we turn immediately to: \mathbf{R}^n ; ℓ_∞ ; $C[a, b]$.

Example 6.3.11.

(i) For $n \in \mathbf{N}$, \mathbf{Q}^n is a countable dense subset of \mathbf{R}^n . This fact can be deduced from the countability of \mathbf{Q} . To show that \mathbf{Q}^n is dense in \mathbf{R}^n , let $B_\varepsilon(x)$ be any open ball of radius $\varepsilon > 0$ and center $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ and use the density of \mathbf{Q} in \mathbf{R} to choose rational numbers r_1, \dots, r_n such that

$$|x_j - r_j| < \varepsilon n^{-2^{-1}} \text{ for } j = 1, \dots, n.$$

Then

$$d_n(x, r) = \left(\sum_{j=1}^n |x_j - r_j|^2 \right)^{2^{-1}} < \varepsilon \text{ and so } r = (r_1, \dots, r_n) \in \mathbf{Q}^n \cap B_\varepsilon(x).$$

(ii) Let ℓ_2 and ℓ_∞ be the spaces of real-valued sequences $(x_k)_{k=1}^\infty$ with $\sum_{j=1}^\infty x_j^2 < \infty$ and $\sup_{k \in \mathbf{N}} |x_k| < \infty$, respectively. For two sequences $x = (x_j)_{j=1}^\infty$ and $y = (y_j)_{j=1}^\infty$, define

$$d_p(x, y) = \begin{cases} (\sum_{j=1}^\infty |x_j - y_j|^2)^{2^{-1}}, & p = 2 \\ \sup_{1 \leq j < \infty} |x_j - y_j|, & p = \infty \end{cases}.$$

Then (ℓ_2, d_2) and (ℓ_∞, d_∞) are metric spaces, but the former is separable and the latter is not separable. It is enough to check the statement after the but. Let S be the set of all sequences of rational numbers that are eventually 0. If $\varepsilon > 0$ is given, then for any $x = (x_j)_{j=1}^\infty \in \ell_2$ there is some $N \in \mathbf{N}$ such that $\sum_{j=N}^\infty |x_j|^2 < (2^{-1}\varepsilon)^2$; at the same time, for each $j \in \{1, \dots, N\}$ there is an $r_j \in \mathbf{Q}$ such that $|r_j - x_j| < \varepsilon(2\sqrt{N})^{-1}$. Accordingly, the sequence $r = (r_1, \dots, r_{N-1}, 0, 0, 0, \dots) \in S$ satisfies

$$d_2(x, r) \leq \left(\sum_{j=1}^{N-1} |x_j - r_j|^2 \right)^{\frac{1}{2}} + \left(\sum_{j=N}^\infty x_j^2 \right)^{\frac{1}{2}} < \varepsilon.$$

This implies that S is dense in ℓ_2 . As for the latter, given $n \in \mathbf{N}$ let

$$x^{(n)} = (x_1^{(n)}, x_2^{(n)}, \dots, x_j^{(n)}, \dots)$$

be an element in ℓ_∞ . Define $x = (x_1, \dots, x_j, \dots)$, where

$$x_j = \begin{cases} x_j^{(j)} + 1, & |x_j^{(j)}| \leq 1 \\ 0, & |x_j^{(j)}| > 1 \end{cases}.$$

Thus $x \in \ell_\infty$. Moreover,

$$d_\infty(x, x^{(n)}) \geq |x_n - x_n^{(n)}| \geq 1 \text{ for all } n \in \mathbf{N}.$$

This shows that the set $\{x^{(n)}\}_{n=1}^\infty$ cannot be dense in ℓ_∞ .

(iii) For any finite closed interval $[a, b] \subset \mathbf{R}$, $C[a, b]$ of all real-valued continuous functions on $[a, b]$ is a metric space under the sup-distance:

$$d(f, g) = \sup_{x \in [a, b]} |f(x) - g(x)| < \infty \text{ for all } f, g \in C[a, b].$$

Moreover, it is separable. In fact, let S be the set of piecewise linear functions of the form

$$f(t) = s_k + \left(\frac{s_{k+1} - s_k}{t_{k+1} - t_k} \right) (t - t_k) \text{ for all } t_k \leq t \leq t_{k+1},$$

where $a = t_0 < t_1 < \dots < t_n = b$ is a partition of $[a, b]$, and the s_k, t_k are rational (with the possible exceptions of $t_0 = a$ and $t_n = b$). The set S is enumerable. Suppose $g \in C[a, b]$. For any $\varepsilon > 0$, there is a $\delta > 0$ such that

$$|t - s| < \delta \Rightarrow |g(t) - g(s)| < 2^{-2}\varepsilon.$$

Let $a = t_0 < t_1 < \dots < t_n = b$ be a partition of $[a, b]$ with t_1, \dots, t_{n-1} rational and such that $\max_{0 \leq k \leq n-1} (t_{k+1} - t_k) < \delta$. Let s_0, \dots, s_n be rational numbers such that $\max_{0 \leq k \leq n} |g(t_k) - s_k| < \frac{\varepsilon}{4}$. Then for $t_k \leq t \leq t_{k+1}$ we have

$$f(t) - g(t) = \left(\frac{t_{k+1} - t}{t_{k+1} - t_k} \right) (s_k - g(t)) + \left(\frac{t - t_k}{t_{k+1} - t_k} \right) (s_{k+1} - g(t)),$$

and hence

$$|f(t) - g(t)| \leq |s_k - g(t_k)| + |g(t_k) - g(t)| + |s_{k+1} - g(t_{k+1})| + |g(t_{k+1}) - g(t)| < \varepsilon.$$

Thus S is dense in $C[a, b]$.

Motivated by the foregoing example, we can establish a relationship between compactness and separability.

Theorem 6.3.12. *Let (X, d) be a metric space. If X is compact, then it is separable.*

Proof. Suppose that $n \in \mathbf{N}$ and X is compact. Since $X = \bigcup_{x \in X} B_{n^{-1}}(x)$, it follows that there is a $j_n \in \mathbf{N}$ such that $X = \bigcup_{j=1}^{j_n} B_{n^{-1}}(x_j)$. Now, the countable set $S = \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{j_n} \{x_j\}$ is dense in X . As a matter of fact, if $x \in X$ then $x \in B_{n^{-1}}(x_k)$ for some $k \in \{1, \dots, j_n\}$. Now for any $\varepsilon > 0$ there exists an $n \in \mathbf{N}$ such that $n^{-1} < \varepsilon$ and hence $x_k \in B_{n^{-1}}(x) \subseteq B_{\varepsilon}(x)$, i.e., $B_{\varepsilon}(x) \cap S \neq \emptyset$. Therefore, X is separable. \square

The next result is Lebesgue's property which is of great generality and combines the concepts of open sets, density and separability. This theorem should be compared to the Heine-Borel property, for each statement makes an assertion about reducing an open covering of a set.

Theorem 6.3.13. *For $n \in \mathbf{N}$ let $X = \mathbf{R}^n$ be equipped with the distance $d_n(\cdot, \cdot)$.*

(i) *There is a countable collection \mathcal{B} of open balls in X such that any open subset of X can be expressed as the union of some subcollection of \mathcal{B} .*

(ii) *If S is any subset of X and $\{A_{\mu}\}_{\mu \in I}$ is a collection of open sets whose union contains S , then there is a countable subcollection $\{A_{\mu_k}\}_{k=1}^{\infty}$ whose union contains S .*

Proof. (i) As a matter of fact, if \mathcal{B} stands for the collection of all open balls $B_r(q)$ with $r > 0$ being a rational number and $q \in \mathbf{Q}^n$, then for a given open

set A we can consider the subcollection \mathcal{B}_A consisting of those open balls such that $B_r(q) \subseteq A$. Their union is obviously contained in A , and the collection is countable because both \mathbf{Q}^n and \mathbf{Q} are countable. Therefore it remains only to show that A is contained in the union of this collection. If x is a point in A , then there is an open ball about x with rational radius r such that $B_r(x) \subseteq A$. Since \mathbf{Q}^n is dense, there is a point q of \mathbf{Q}^n in $B_{2^{-1}r}(x)$. Then x is in $B_{2^{-1}r}(q)$, and $B_{2^{-1}r}(q)$ is in \mathcal{B}_A because it is contained in $B_r(x)$, which lies in A . This last assertion is verified by observing that

$$y \in B_{2^{-1}r}(q) \Rightarrow d_n(x, y) \leq d_n(x, q) + d_n(q, y) < r.$$

We have shown that an arbitrary point $x \in A$ is contained in some open ball in \mathcal{B}_A , so A is contained in the union of all balls in \mathcal{B}_A .

(ii) Suppose $S \subseteq \bigcup_{\mu \in I} A_\mu$. By (i), each of the open sets A_μ can be written as the union of all the balls $B_{r_k}^\mu$ centered at the points of \mathbf{Q}^n that have rational radii and are contained in A_μ , say, $A_\mu = \bigcup_{k=1}^\infty B_{r_k}^\mu$. If $x \in S$ then there is some r_k such that $x \in B_{r_k}^\mu \subseteq A_\mu$. For each of the open balls $\{B_{r_k}^\mu\}$, choose just one A_μ that contains $B_{r_k}^\mu$ and call it $A_{r_k}^\mu$. The subcollection of all such $A_{r_k}^\mu$ is countable because there is only one $A_{r_k}^\mu$ for each $B_{r_k}^\mu$. Also, the union of this subcollection contains the union of the $B_{r_k}^\mu$'s, which in turn contains S . Hence there is a countable subcollection of $\{A_\mu\}_{\mu \in I}$ whose union contains S . \square

Remark 6.3.14. *It is worth remarking that the hypothesis of Theorem 6.3.13 (ii) makes no restrictive assumption about D . That is the strength and generality of this result; it applies to any subset D of \mathbf{R}^n . If it is given, in addition, that D is closed and bounded, then the countable subcollection obtained from Lebesgue's property can be further reduced to a finite subcollection whose union still contains D .*

Definition 6.3.15. *Suppose that (X, d) is a metric space.*

- (i) *A set $S \subseteq X$ is said to be nowhere dense if \overline{S} contains no open ball, i.e., $\overline{S}^\circ = \emptyset$.*
- (ii) *A set $S \subseteq X$ is said to be of first category in X if S is a countable union of nowhere dense sets.*
- (iii) *A set $S \subseteq X$ is said to be of second category in X if S is not of first category.*

Example 6.3.16.

(i) *Cantor's ternary set C is nowhere dense. In fact, $m_{id}(C) = 0$ implies that if $(\overline{C})^\circ = C^\circ$ were not empty, then C would contain a finite open interval (a, b) , then $b - a \leq m_{id}(C) = 0$, and hence $a = b$, a contradiction. Thus, $C^\circ = \emptyset$.*

(ii) *\mathbf{Q} is of first category in \mathbf{R} – this follows from the definition.*

(iii) *\mathbf{R} is of second category in itself – this assertion is not obvious, and in fact it can only be deduced as a consequence of the so-called Baire category theorem (named after René-Louis Baire) as follows.*

Theorem 6.3.17. *Given a metric (X, d) . If X is complete, then it is of second category, and hence the complement of a first category set in X is of second category and dense in X*

Proof. Let A_k be nowhere dense in X for every $k \in \mathbf{N}$, and let $E = \bigcup_{k=1}^{\infty} A_k$. We shall show that there is $x \in X \setminus E = E^c$ and this will establish the result. Let $x^{(0)} \in X$. Since A_1 is nowhere dense, there is a closed ball B_1 of radius less than 2^{-1} inside the closed ball $\overline{B}_1(x^{(0)})$ and such that $B_1 \cap A_1 = \emptyset$. Since A_2 is nowhere dense, there is also a closed ball B_2 of radius smaller than $1/3$ such that $B_2 \subseteq B_1$ and $B_2 \cap A_2 = \emptyset$. Continuing this construction produces a nested sequence of closed balls: $\{B_k\}_{k=1}^{\infty}$ such that the radius of B_k is less than $(k+1)^{-1}$. By the proof for Theorem 6.2.10 and its immediate Remark 6.2.11, there is $x \in \bigcap_{k=1}^{\infty} B_k \subseteq B_1(x^{(0)})$. But $x \notin A_k$ for any k ; so $x \notin E$.

If S is of first category, then the density of S^c can be worked out by replacing $B_1(x^{(0)})$ by $B_\epsilon(x^{(0)})$ in the foregoing argument. Of course, S^c must be second category: Otherwise, $X = S \cup S^c$ is of first category, a contradiction. \square

Problems

6.1. *For each of the following functions on \mathbf{R}^2 , determine which of the properties in Definition 6.1.1 (i)-(ii)-(iii) are satisfied:*

(i) $d(x, y) = |x_1 - y_1| + |x_2 - y_2|$;

(ii) $d(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$;

(iii)

$$d(x, y) = \begin{cases} ((x_1 - y_1)^2 + (x_2 - y_2)^2)^{2^{-1}}, & x_2 \geq y_2 \\ |x_1 - y_1| + |x_2 - y_2|, & x_2 \leq y_2 \end{cases}.$$

6.2. Prove that if $x^{(1)}, \dots, x^{(m)}$ are points in a metric space (X, d) , then $d(x^{(1)}, x^{(m)}) \leq \sum_{j=2}^m d(x^{(j)}, x^{(j-1)})$.

6.3. Given $n \in \mathbf{N}$.

(i) Let $d_n^*(x, y) = \sum_{k=1}^n |x_k - y_k|$ for $x, y \in \mathbf{R}^n$. Prove that d_n^* satisfies Definition 6.1.1(i)-(ii)-(iii). In addition, show that $d_n(x, y) \leq d_n^*(x, y)$ for $x, y \in \mathbf{R}^n$.

(ii) Let $d_{*,n}(x, y) = \max_{1 \leq k \leq n} |x_k - y_k|$ for $x, y \in \mathbf{R}^n$. Prove that $d_{*,n}$ satisfies Definition 6.1.1(i)-(ii)-(iii). Discuss the relationship between d_n and $d_{*,n}$.

6.4. Prove that A° and \bar{A} are the largest open subset of A and the smallest closed set containing A , respectively.

6.5. In \mathbf{R}^2 , show that the closure of $B_1(0)$ is $\bar{B}_1(0)$.

6.6. Suppose (X, d) is a metric space. Prove that if S is a finite subset of X , then S is disconnected.

6.7. Prove that, if $f, g > 0$ on \mathbf{R} so that their graphs

$$G_1 = \{x \in \mathbf{R}^2 : x_2 = f(x_1)\} \text{ and } G_2 = \{x \in \mathbf{R}^2 : x_2 = -g(x_1)\}$$

are sets in \mathbf{R}^2 , then $S = G_1 \cup G_2$ is disconnected.

6.8. Decide whether or not each of the following sequences in \mathbf{R}^2 :

(i) $x^{(k)} = ((k-1)k^{-1}, (k+1)k^{-1})$;

(ii) $x^{(k)} = (k^{-1}, \sin \pi k)$;

(iii) $x^{(k)} = (k^{-1}, k)$;

converges and, if so, find its limit.

6.9. Let X be the interval $(0, 1)$ in \mathbf{R} with the usual distance $|x - y|$. Find a Cauchy sequence in X that does not converge in X .

6.10. For $n \in \mathbf{N}$, prove the following two equivalent statements:

(i) $\{x^{(k)}\}_{k=1}^\infty$ is a bounded sequence in \mathbf{R}^n if and only if each coordinate sequence $\{x_j^{(k)}\}_{k=1}^\infty$ ($j = 1, \dots, n$) is bounded in \mathbf{R} ;

(ii) A is a bounded set in \mathbf{R}^n when and only when there exists an n -cube containing A .

6.11. Prove that if X is an infinite set and its distance function d is given by

$$d(x, y) = \begin{cases} 1, & x \neq y \\ 0, & x = y \end{cases},$$

then every Cauchy sequence in X is eventually constant – there exists an $N \in \mathbf{N}$ such that $x^{(k)} = x^{(N)}$ whenever $k \geq N$.

6.12. For $n \in \mathbf{N}$, prove that $D = \{x = (x_1, \dots, x_n) \in \mathbf{R}^n : x_1, \dots, x_n \in \mathbf{R} \setminus \mathbf{Q}\}$ is dense in \mathbf{R}^n .

6.13. Let X be a metric space with the distance function d . Show that if $\{x^{(k)}\}_{k=1}^\infty$ is a Cauchy sequence in X and has a subsequence $\{x^{(k_j)}\}_{j=1}^\infty$ that converges to a point $x \in X$, then $d(x_k, x) \rightarrow 0$.

6.14. Let (X, d) be a metric space. Prove that X is complete if and only if every closed nested sequence of closed balls $\{\overline{B}_{r_k}(x_k)\}_{k=1}^\infty$ with $r_k \rightarrow 0$ has non-void intersection.

6.15. Let (X, d) be a complete metric space. Prove that if $\{S_j\}_{j=1}^\infty$ is a sequence of nonempty closed subsets of X with $S_{j+1} \subseteq S_j$ for any $j \in \mathbf{N}$ and $\text{diam}(S_j) = \sup\{d(x, y) : x, y \in S_j\} \rightarrow 0$ as $j \rightarrow \infty$ then $\bigcap_{j=1}^\infty S_j \neq \emptyset$.

6.16. Suppose that (X, d) is a metric space. Show that $S \subseteq X$ is nowhere dense if and only if every open ball O contains an open ball O' such that $O' \cap S = \emptyset$.

6.17. Suppose that (X, d) is a metric space. Prove that if $E_\alpha \subseteq X$ is compact for any $\alpha \in I$, then $\bigcap_{\alpha \in I} E_\alpha$ is compact.

6.18. Let (X, d) be a metric space and $\lim_{n \rightarrow \infty} d(x_n, x_0) = 0$. Prove that $\{x_j\}_{j=0}^\infty$ is compact.

Chapter 7

Continuous Mappings

Since every metric space is automatically a topological space with the set of all open sets as the metrizable topology, we have a notion of continuous mappings between metric spaces. Note that a function between topological spaces is said to be continuous if the inverse image of every open set is open – this is an attempt to capture the intuition that there are no breaks or separations in the function. Without referring to the topology, this notion can also be directly defined using limits of sequences. In this chapter, we consider properties of continuous mappings, contractions with straightforward applications to integral and differential equations, and equivalence of metric spaces.

7.1 Criteria for Continuity

By a mapping on a metric space (X, d) , frequently denoted (X, d_X) in what follows, we mean a mapping on the set of points in X , and by a mapping with values in a metric space Y we mean a mapping with values in the set of points in Y . Thus if $f : X \rightarrow Y$ is a mapping from the metric space X into the metric space Y , then to each point $x \in X$ is associated a point $f(x) \in Y$. The mapping f will be called continuous at a point $x_0 \in X$ if, roughly speaking, points of X that are near x_0 are mapped by f into points of Y that are near $f(x_0)$. Below is the precise definition.

Definition 7.1.1. *Let (X, d_X) and (Y, d_Y) be two metric spaces, let $f : X \rightarrow Y$ be a mapping, and let $x_0 \in X$. Then f is said to be continuous at x_0 if, given any*

$\varepsilon > 0$, there exists a $\delta > 0$ such that

$$x \in X \text{ and } d_X(x, x_0) < \delta \Rightarrow d_Y(f(x), f(x_0)) < \varepsilon.$$

Moreover, if f is continuous at all points in X , then f is said to be continuous on X .

Here we want to emphasize that δ depends on ε (as well as x_0). Therefore, we could more accurately have written $\delta(\varepsilon)$ instead of δ . We stick to the notation δ rather than $\delta(\varepsilon)$ for notational simplicity, always bearing in mind that each ε must have its own δ .

Example 7.1.2.

(i) The mapping $f : \mathbf{R} \rightarrow \mathbf{R}$ given by $f(x) = x^2$ is continuous.

(ii) Let X be any metric space with the distance d , and x_0 a fixed point in X . Then the mapping $f : X \rightarrow \mathbf{R}$ given by $f(x) = d(x, x_0)$ for all $x \in X$ is continuous. This follows from a use of the triangle inequality of d . The special case $X = \mathbf{R}$ and $x_0 = 0$ shows that the absolute value function $|x|$ is continuous.

(iii) If $f : X \rightarrow Y$ is continuous and S is a subspace of X , then the restriction of f to S is continuous on S . This is clear from Definition 7.1.1. One of the curious consequences of Definition 7.1.1 occurs at a point in $x_0 \in S$ that is not a cluster point of S . Since x_0 is an isolated point of S , there is an open ball $B_\delta(x_0)$ that contains no point of S other than x_0 . Accordingly, the mapping is always continuous at x_0 .

Let us have a look at another example.

Example 7.1.3. For $n, m \in \mathbf{N}$, let $T : \mathbf{R}^n \rightarrow \mathbf{R}^m$ be a linear mapping; that is,

$$T(ax + by) = aT(x) + bT(y) \text{ for all } x, y \in \mathbf{R}^n.$$

Then there exists an $m \times n$ real-valued matrix $A = [a_{kj}]$ such that $y = T(x) = Ax$, i.e.,

$$T \left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right) = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},$$

where

$$y_j = \sum_{k=1}^n a_{jk}x_k \text{ with } a_{kj} \in \mathbf{R}; \quad k = 1, \dots, n; \quad j = 1, \dots, m.$$

Thus T is continuous.

To prove this result, we first consider the case $m = 1$, and see that $T : \mathbf{R}^n \rightarrow \mathbf{R}$ is a function. Let $e^{(k)}$, $k = 1, \dots, n$ be the standard basis vectors of \mathbf{R}^n : Its k -th coordinate is 1 and others are 0. And set $A_k = T(e^{(k)})$. Since $x = \sum_{k=1}^n x_k e^{(k)}$, by the definition of T we have

$$T(x) = \sum_{k=1}^n x_k T(e^{(k)}) = \sum_{k=1}^n x_k A_k.$$

Thus T is represented by the row matrix $A = [A_1, \dots, A_n]$ whenever x is written as the column matrix

$$x = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$$

Given $m > 1$. Then

$$T(x) = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_m \end{pmatrix},$$

where y_j is a function of x , say $y_j = f_j(x)$. The linearity of T implies that T is linear in each coordinate; that is, each f_j is a linear function from \mathbf{R}^n into \mathbf{R} . Therefore, by the first case, each f_j is given by a row matrix:

$$y_j = f_j(x) = \sum_{k=1}^n a_{jk}x_k.$$

This set of m linear functions f_1, \dots, f_m determines the m rows of the matrix $[a_{jk}]$. Thus $T(x) = Ax$.

In order to prove the continuity of T , for

$$x = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_l \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_l \end{pmatrix} \quad \text{in } \mathbf{R}^l$$

let

$$\begin{cases} M_T = (\sum_{j=1}^m \sum_{k=1}^n a_{jk}^2)^{2^{-1}}; \\ \|x\|_l = (\sum_{k=1}^l x_k^2)^{2^{-1}}; \\ d_l(x, y) = \|x - y\|_l. \end{cases}$$

Then

$$\|T(x)\|_m = \left(\left(\sum_{k=1}^n a_{1k} x_k \right)^2 + \dots + \left(\sum_{k=1}^n a_{mk} x_k \right)^2 \right)^{2^{-1}}.$$

Using the Cauchy-Schwarz inequality, we get

$$\left(\sum_{k=1}^n a_{jk} x_k \right)^2 \leq \left(\sum_{k=1}^n a_{jk}^2 \right) \left(\sum_{k=1}^n x_k^2 \right) \quad \text{for } j = 1, 2, \dots, m.$$

and hence $\|T(x)\|_m \leq M_T \|x\|_n$. Consequently, for any fixed point $x_0 \in \mathbf{R}^n$ we have

$$d_m(T(x), T(x_0)) = \|T(x) - T(x_0)\|_m \leq M_T d_n(x, x_0).$$

If $\varepsilon > 0$ is given, then we can define $\delta = \varepsilon(1 + M_T)^{-1}$, and hence

$$d_m(T(x), T(x_0)) < \varepsilon \quad \text{whenever } d_n(x, x_0) < \delta.$$

Remark 7.1.4. Note that the choice of δ above does not depend on x_0 where the continuity is established. Thus we conclude that linear mappings, like linear functions on \mathbf{R} , are uniformly continuous on \mathbf{R}^n .

Definition 7.1.1 may be reformulated by saying that f is continuous at x_0 if, given any open ball in Y of center $f(x_0)$, there exists an open ball in X centered at x_0 whose image under f is contained in the former ball. Clearly, these open balls can be replaced by open sets. The forthcoming criterion for the continuity of a mapping from one metric space into another is often useful.

Theorem 7.1.5. *Let (X, d_X) and (Y, d_Y) be two metric spaces and let $f : X \rightarrow Y$ be a mapping. Then f is continuous if and only if for every open subset V of Y the inverse image $f^{-1}(V) = \{x \in X : f(x) \in V\}$ is an open subset of X .*

Proof. To prove this theorem, first suppose that f is continuous. We have to show that if $V \subseteq Y$ is open, then also $f^{-1}(V)$ is open. Let $x_0 \in f^{-1}(V)$. Then $f(x_0) \in V$. Since V is open, it contains the open ball in Y of center $f(x_0)$ and some radius $\epsilon > 0$. Since f is continuous at x_0 , there is a $\delta > 0$ such that if $x \in X$ and $d_X(x, x_0) < \delta$ then $d_Y(f(x), f(x_0)) < \epsilon$. This means that if x is in the open ball in X of center x_0 and radius δ then $f(x)$ belongs to the open ball in Y of center $f(x_0)$ and radius ϵ , so that $f(x) \in V$. That is, $f^{-1}(V)$ contains the open ball in X of center x_0 and radius δ . Since x_0 was any point of $f^{-1}(V)$, the set $f^{-1}(V)$ is open.

Conversely, suppose that for every open $V \subset Y$, the set $f^{-1}(V)$ is an open subset of X . We must show that f is continuous at any point $x_0 \in X$. For any $\epsilon > 0$ let $B_\epsilon(f(x_0)) \subseteq Y$ be the open ball of center $f(x_0)$ and radius ϵ . Then $f^{-1}(B_\epsilon(f(x_0)))$ is an open subset of X that contains x_0 , and hence contains the open ball in X of center x_0 and some radius $\delta > 0$. Thus

$$x \in X \text{ and } d_X(x, x_0) < \delta \Rightarrow d_Y(f(x), f(x_0)) < \epsilon.$$

Namely, f is continuous at x_0 . □

Remark 7.1.6. *As an immediate consequence of this theorem, we have that if $f : X \rightarrow \mathbf{R}$ is continuous then for any $c \in \mathbf{R}$*

$$\{x \in X : f(x) > c\} \text{ and } \{x \in X : f(x) < c\}$$

are open subsets of X . Additionally, it is not hard to see that a continuous mapping of a continuous mapping is a continuous mapping; that is, if $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ are continuous, so is the mapping $g \circ f : X \rightarrow Z$.

The next result is the sequence criterion for continuity.

Theorem 7.1.7. *Let (X, d_X) and (Y, d_Y) be two metric spaces and let $f : X \rightarrow Y$ be a mapping. Then f is continuous if and only if for every sequence $\{x_k\}_{k=1}^\infty$ of points in X that converges to $x_0 \in X$, the sequence $\{f(x_k)\}_{k=1}^\infty$ converges to $f(x_0)$ in Y .*

Proof. Suppose first that f is continuous at $x_0 \in X$ and that $\{x_k\}_{k=1}^\infty$ converges to x_0 in X . We have to show that $\{f(x_k)\}_{k=1}^\infty$ converges to $f(x_0)$ in Y . Given $\varepsilon > 0$, the continuity of f at x_0 implies that there is a $\delta > 0$ such that $d_Y(f(x), f(x_0)) < \varepsilon$ whenever $x \in X$ and $d_X(x, x_0) < \delta$. Because $\{x_k\}_{k=1}^\infty$ converges to x_0 , there exists a positive integer n_0 such that $d_X(x_k, x_0) < \delta$ for all $k > n_0$. Hence if $k > n_0$ then $d_Y(f(x_k), f(x_0)) < \varepsilon$, which shows that $\{f(x_k)\}_{k=1}^\infty$ converges to $f(x_0)$.

We now suppose that the statement after the if and only if is valid. If f is not continuous at x_0 , then there is some $\varepsilon_0 > 0$ such that for no number $\delta > 0$ it is true that whenever $x \in X$ and $d_X(x, x_0) < \delta$ then necessarily $d_Y(f(x), f(x_0)) < \varepsilon$. Hence for any $k \in \mathbf{N}$ we can find a point $x_k \in X$ such that $d_X(x_k, x_0) < k^{-1}$ and $d_Y(f(x_k), f(x_0)) \geq \varepsilon_0$. Since $d_X(x_k, x_0) < k^{-1}$, the sequence $\{x_k\}_{k=1}^\infty$ converges to x_0 . However $\{f(x_k)\}_{k=1}^\infty$ does not converge to $f(x_0)$ since $d_Y(f(x_k), f(x_0)) \geq \varepsilon_0$ for all k . This completes the proof. \square

Example 7.1.8. Let f be defined on \mathbf{R}^2 by

$$f(x) = f(x_1, x_2) = \begin{cases} 0 & , x = 0 \\ x_1 x_2 (x_1^2 + x_2^2)^{-1} & , x \neq 0 \end{cases}.$$

Then as $\{x^{(k)}\}_{k=1}^\infty$ approaches 0 along the axes, we see that $\{f(x^{(k)})\}_{k=1}^\infty$ converges to 0. But if $x^{(k)} = (k^{-1}, k^{-1})$, then $f(x^{(k)}) = 2^{-1}$. Hence f is not continuous at 0.

7.2 Continuous Mappings over Compact or Connected Metric Spaces

First, we consider continuous mappings on a compact metric space.

Theorem 7.2.1. Let (X, d_X) and (Y, d_Y) be two metric spaces and let $f : X \rightarrow Y$ be a continuous mapping. If X is compact, so is its image $f(X)$.

Proof. We must prove that if $f(X) = \{f(x) : x \in X\}$ is covered by the union of a collection of open subsets of Y then it is covered by the union of a finite number of these open subsets. So, suppose that $f(X) \subseteq \bigcup_{j \in I} V_j$ where I is an index set and V_j is open. Because f is continuous, each inverse image $f^{-1}(V_j)$ is open. Also, for any $x \in X$ we have $f(x) \in V_j$ for some $j \in I$, in which case

$x \in f^{-1}(V_j)$, so that $X = \cup_{j \in I} f^{-1}(V_j)$. By the compactness of X we can find a finite subset $J \subseteq I$ such that $X = \cup_{j \in J} f^{-1}(V_j)$. Therefore

$$f(X) = \cup_{j \in J} f(f^{-1}(V_j)) \subseteq \cup_{j \in J} V_j.$$

Thus $f(X)$ is compact. □

This theorem has two extremely important immediate consequences.

Corollary 7.2.2. *Let (X, d_X) and (Y, d_Y) be two metric spaces and let $f : X \rightarrow Y$ be a continuous mapping. If X is compact, then f is bounded.*

Proof. Since $f(X)$ is compact, it is bounded according to Corollary 6.3.9. □

The example $f(x) = x$ on $(0, 1)$ shows that a bounded continuous real-valued function need not assume either its least upper bound or its greatest lower bound on its domain of definition. If, however, the domain is compact, then these values are actually taken on by the continuous function.

Corollary 7.2.3. *Let (X, d_X) be a nonempty compact metric space and $f : X \rightarrow \mathbf{R}$ be continuous. Then*

- (i) $\sup\{f(x) : x \in X\} = M$ and $\inf\{f(x) : x \in X\} = m$ are finite.
- (ii) There are points $x, y \in X$ such that $f(x) = M$ and $f(y) = m$.

Proof. (i) By Theorem 7.2.1 it follows that $f(X)$ is a compact subset of \mathbf{R} , hence closed and bounded. Since $f(X)$ is not empty, $f(X)$ has both a greatest and a least element.

To prove (ii), by the definition of M we derive that for each $k \in \mathbf{N}$ there is $x_k \in X$ such that $M - k^{-1} < f(x_k) \leq M$. Since X is compact, $\{x_k\}_{k=1}^{\infty}$ has a convergent subsequence $\{x_{k_j}\}_{j=1}^{\infty}$, which converges to a point $x \in X$. Since f is continuous, $f(x) = \lim_{j \rightarrow \infty} f(x_{k_j}) = M$. The argument for the existence of the point y such that $f(y) = m$ is similar. □

Definition 7.2.4. *Let (X, d_X) and (Y, d_Y) be two metric spaces and let $f : X \rightarrow Y$ be a mapping. Then f is said to be uniformly continuous if, given any $\epsilon > 0$, there exists a $\delta > 0$ such that if $x, y \in X$ and $d_X(x, y) < \delta$ then $d_Y(f(x), f(y)) < \epsilon$.*

If it happens that a mapping $f : X \rightarrow Y$ is such that for a certain subset S of X , the restriction of f to S is uniformly continuous, we say that f is uniformly continuous on S .

Clearly, a uniformly continuous mapping is continuous. To check continuity at a point $x_0 \in X$ just set $y = x_0$ in the definition. The next theorem will state that conversely if f is continuous then f is actually uniformly continuous, provided X is compact.

Theorem 7.2.5. *Let (X, d_X) and (Y, d_Y) be two metric spaces and let $f : X \rightarrow Y$ be a continuous mapping. If X is compact, then f is uniformly continuous; equivalently, $\{f(x_j)\}_{j=1}^\infty$ is a Cauchy sequence in Y whenever $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence in X .*

Proof. Suppose f is not uniformly continuous on X . Then there are an $\varepsilon_0 > 0$ and sequences $\{x_k\}_{k=1}^\infty$ and $\{y_k\}_{k=1}^\infty$ of points in X such that

$$d_X(x_k, y_k) < k^{-1} \quad \text{and} \quad d_Y(f(x_k), f(y_k)) \geq \varepsilon_0.$$

Since X is compact, $\{x_k\}_{k=1}^\infty$ has a subsequence $\{x_{k_j}\}_{j=1}^\infty$, which converges to a point $x \in X$. Then $y_{k_j} \rightarrow x$ since

$$0 \leq d_X(y_{k_j}, x) \leq d_X(y_{k_j}, x_{k_j}) + d_X(x_{k_j}, x) < k_j^{-1} + d_X(x_{k_j}, x).$$

Since f is continuous and

$$d_Y(f(x_{k_j}), f(y_{k_j})) \leq d_Y(f(x_{k_j}), f(x)) + d_Y(f(y_{k_j}), f(x)),$$

we have

$$\begin{cases} \lim_{j \rightarrow \infty} d_Y(f(x_{k_j}), f(x)) = 0; \\ \lim_{j \rightarrow \infty} d_Y(f(y_{k_j}), f(x)) = 0; \\ \lim_{j \rightarrow \infty} d_Y(f(x_{k_j}), f(y_{k_j})) = 0, \end{cases}$$

contradicting $d_Y(f(x_{k_j}), f(y_{k_j})) \geq \varepsilon_0$.

Next, we prove the equivalence. The above argument indicates that if f is not uniformly continuous on X , then the sequence $\{x_{k_1}, y_{k_1}, x_{k_2}, y_{k_2}, \dots\}$ is a Cauchy sequence in X but $\{f(x_{k_1}), f(y_{k_1}), f(x_{k_2}), f(y_{k_2}), \dots\}$ is not a Cauchy sequence in Y . On the other hand, if f is uniformly continuous, then we conclude that for any $\varepsilon > 0$ there is a $\delta > 0$ such that $d_X(x, y) < \delta$ implies $d_Y(f(x), f(y)) < \varepsilon$. Now if $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence in X then for the above obtained $\delta > 0$ there exists $n_0 > 0$ such that $d_X(x_j, x_k) < \delta$ as $k, j > n_0$ and consequently, $d_Y(f(x_j), f(x_k)) < \varepsilon$, i.e., $\{f(x_j)\}_{j=1}^\infty$ is a Cauchy sequence in Y . \square

Remark 7.2.6. *The compactness of the domain is important. For example, the function f , defined by $f(x) = x^{-1}$ on $(0, 1]$, is continuous but is not uniformly continuous on the noncompact set $(0, 1]$. The problem, of course, lies at the left end-point of the interval. The function values are arbitrarily large in every open ball of center 0.*

Secondly, we discuss continuous mappings on a connected metric space.

Theorem 7.2.7. *Let (X, d_X) and (Y, d_Y) be two metric spaces and let $f : X \rightarrow Y$ be a continuous mapping. If X is connected, so is its image $f(X)$.*

Proof. If $f(X)$ is not connected, then $f(X) = A \cup B$, where A and B are disjoint nonempty open subsets of Y . It is clear that $f^{-1}(A)$ and $f^{-1}(B)$ are disjoint nonempty open subsets of X . We therefore have the expression of X as $X = f^{-1}(A) \cup f^{-1}(B)$. This contradicts the fact that X is connected, proving the theorem. \square

Below is the intermediate value property as a consequence of the above theorem.

Example 7.2.8. *Suppose $[a, b]$ is a finite interval of \mathbf{R} . Let $f : [a, b] \rightarrow \mathbf{R}$ be continuous with $f(a) < f(b)$. If $f(a) < y < f(b)$, then there is $\zeta \in [a, b]$ such that $f(\zeta) = y$.*

In fact, since $[a, b]$ is connected, so is $f([a, b])$. Note that any open subset of \mathbf{R} which contains a and b and does not contain some point between a and b is not connected: for suppose that $a < c < b$ and S is an open subset of \mathbf{R} with $a, b \in S$, $c \notin S$, then

$$S = (S \cap \{x \in \mathbf{R} : x < c\}) \cup (S \cap \{x \in \mathbf{R} : x > c\})$$

expresses S as the union of two disjoint nonempty open subsets. In other words, any connected subset of \mathbf{R} contains all points between any two of its points. Since y is between $f(a)$ and $f(b)$ in $f([a, b])$ which is connected, we therefore have $y \in f([a, b])$.

7.3 Sequences of Mappings

Definition 7.3.1. *Let (X, d_X) and (Y, d_Y) be two metric spaces, and for $k \in \mathbf{N}$ let $f_k : X \rightarrow Y$ be a mapping. We say that:*

- (i) $\{f_k\}_{k=1}^\infty$ converges at $x \in X$ provided that $\{f_k(x)\}_{k=1}^\infty$ converges in Y ;
- (ii) $\{f_k\}_{k=1}^\infty$ converges on X , provided that $\{f_k(x)\}_{k=1}^\infty$ converges at any $x \in X$;
- (iii) $\{f_k\}_{k=1}^\infty$ converges to a mapping $f : X \rightarrow Y$, denoted $f = \lim_{k \rightarrow \infty} f_k$, provided that $f(x) = \lim_{k \rightarrow \infty} f_k(x)$ whenever $x \in X$.

Example 7.3.2.

- (i) Let $f_k : [0, 1] \rightarrow \mathbf{R}$ be given by $f_k(x) = x - k^{-1}x$. Then $\lim_{k \rightarrow \infty} f_k(x) = x$.
- (ii) Let $f_k : [0, 1] \rightarrow \mathbf{R}$ be given by $f_k(x) = x^k$. Then

$$\lim_{k \rightarrow \infty} f_k(x) = f(x) = \begin{cases} 0, & x \in [0, 1] \\ 1, & x = 1 \end{cases}.$$

Note that f_k is continuous, but f is not.

Definition 7.3.3. Let (X, d_X) and (Y, d_Y) be two metric spaces, for $k \in \mathbf{N}$ let $f_k : X \rightarrow Y$ be a mapping, and let $f : X \rightarrow Y$ be another mapping. Then we say that $\{f_k\}_{k=1}^\infty$ converges uniformly to f on X if, given any $\varepsilon > 0$, there is an $n_0 \in \mathbf{N}$ such that $d_Y(f_k(x), f(x)) < \varepsilon$ whenever $k > n_0$ for all $x \in X$. If the restrictions of $\{f_k\}_{k=1}^\infty$ to a certain subset S of X converge uniformly to some mapping on S , we say that $\{f_k\}_{k=1}^\infty$ converges uniformly on S .

Uniform convergence clearly implies convergence. Having a look at Example 7.3.2, we have that the first one is uniform, but not the second. Moreover, we have the following Cauchy criterion for uniform convergence.

Theorem 7.3.4. Let (X, d_X) and (Y, d_Y) be two metric spaces, let Y be complete, and for $k \in \mathbf{N}$ let $f_k : X \rightarrow Y$ be a mapping. Then $\{f_k\}_{k=1}^\infty$ converges uniformly to f on X when and only when for any $\varepsilon > 0$ there is an $n_0 \in \mathbf{N}$ such that $d_Y(f_k(x), f_l(x)) < \varepsilon$ whenever $k, l > n_0$ for all $x \in X$.

Proof. It is enough to verify the sufficiency. Suppose that the statement after the when and only when is true. So, for any $x \in X$, $\{f_k(x)\}_{k=1}^\infty$ is a Cauchy sequence in Y . Since Y is complete, this sequence has a limit. Thus $\{f_k\}_{k=1}^\infty$ converges to some mapping f . Given $\varepsilon > 0$, choose an integer n_0 so that we have $d_Y(f_k(x), f_l(x)) < 2^{-1}\varepsilon$ whenever $k, l > n_0$, for all $x \in X$. Then for any fixed $k > n_0$ and fixed $x \in X$ the sequence $\{f_l(x)\}_{l=1}^\infty$ is such that all terms after the n_0 -th are within distance $2^{-1}\varepsilon$ of $f_k(x)$, and are therefore in the closed ball in Y of center $f_k(x)$ and radius $2^{-1}\varepsilon$. Since closed balls are closed sets, the limit

$f(x)$ of the convergent sequence $\{f_l(x)\}_{l=1}^{\infty}$ is also in this closed ball, so that $d_Y(f_k(x), f(x)) \leq 2^{-1}\varepsilon$. Hence if $k > n_0$ we have $d_Y(f_k(x), f(x)) < \varepsilon$ for all $x \in X$, proving uniform convergence. \square

Interestingly, the continuity is preserved under the uniform convergence.

Theorem 7.3.5. *Let (X, d_X) and (Y, d_Y) be two metric spaces and let $\{f_k\}_{k=1}^{\infty}$ be a uniformly convergent sequence of continuous mappings from X into Y . Then $f = \lim_{k \rightarrow \infty} f_k$ is continuous.*

Proof. Fix $x_0 \in X$. Let $\varepsilon > 0$. Take an $n_0 \in \mathbf{N}$ such that $d_Y(f_k(x), f(x)) < 3^{-1}\varepsilon$ for all $x \in X$, which is possible by the uniform convergence. Since each f_k is continuous at x_0 , there is a $\delta > 0$ such that if $x \in X$ and $d_X(x, x_0) < \delta$ then $d_Y(f_k(x), f_k(x_0)) < 3^{-1}\varepsilon$. Hence if $x \in X$ and $d_X(x, x_0) < \delta$ we have

$$d_Y(f(x), f(x_0)) \leq d_Y(f(x), f_k(x)) + d_Y(f_k(x), f_k(x_0)) + d_Y(f_k(x_0), f(x_0)) < \varepsilon.$$

Thus f is continuous at x_0 . We are done. \square

If f_1 and f_2 are mappings from X into Y , it is natural to try to find some measure of the extent to which f_1 and f_2 differ, that is to find some sort of distance between f_1 and f_2 . For any specific $x \in X$ we may say that f_1 and f_2 differ at x by the distance between their values at x , that is by $d_Y(f_1(x), f_2(x))$, but we would really like to measure how much f_1 and f_2 differ over all points of X , not just at x . There are various ways to do this, depending on the circumstances and purposes in mind, but the most simple-minded method turns out to be one of the most useful. It is to take the distance between f_1 and f_2 to be $\max \{d_Y(f_1(x), f_2(x)) : x \in X\}$ if this maximum happens to exist. In order to develop this idea we turn aside for a simple lemma.

Lemma 7.3.6. *Let (X, d_X) and (Y, d_Y) be two metric spaces and let f_1 and f_2 be continuous mappings from X into Y . Then the function $x \mapsto d_Y(f_1(x), f_2(x))$ is continuous on X .*

Proof. This follows from the triangle inequality based estimate below:

$$\begin{aligned} & |d_Y(f_1(x), f_2(x)) - d_Y(f_1(x_0), f_2(x_0))| \\ & \leq |d_Y(f_1(x), f_2(x)) - d_Y(f_1(x), f_2(x_0))| + |d_Y(f_1(x), f_2(x_0)) - d_Y(f_1(x_0), f_2(x_0))| \\ & \leq d_Y(f_2(x), f_2(x_0)) + d_Y(f_1(x), f_1(x_0)). \end{aligned}$$

\square

Now we consider the set $C(X, Y)$ of all continuous mappings from X into Y . If X is compact, then it is clear that for $f_1, f_2 \in C(X, Y)$ the following notion

$$d_{C(X,Y)}(f_1, f_2) = \max \{d_Y(f_1(x), f_2(x)) : x \in X\}$$

makes sense since any continuous real-valued function on a compact metric space attains a maximum. Moreover, it is easy to prove that $C(X, Y)$ is a metric space under $d_{C(X,Y)}(\cdot, \cdot)$. However, it is abstract in the sense that its points are mappings on another metric space. A sequence of points in $C(X, Y)$ is a sequence of continuous mappings $\{f_k\}_{k=1}^\infty$ from X into Y . This sequence will converge to a point $f \in C(X, Y)$ if and only if $\lim_{k \rightarrow \infty} d_{C(X,Y)}(f_k, f) = 0$; in other words, if and only if for each $\varepsilon > 0$ there is an $n_0 \in \mathbf{N}$ such that for any integer $k > n_0$ one has $d_{C(X,Y)}(f_k, f) < \varepsilon$; that is, $d_Y(f_k(x), f(x)) < \varepsilon$ for all $x \in X$. Equivalently, $\{f_k\}_{k=1}^\infty$ converges uniformly to f .

An application of the last two theorems gives the following result covering the well-known Arzelà-Ascoli theorem which comes about as this is a generalization going back to Cesare Arzelà's theorem – if f_n is a uniformly bounded sequence of Riemann integrable functions that converge pointwise to a Riemann integrable function f on the finite interval $[a, b]$, then

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} \int_a^b f_n(x)dx.$$

Theorem 7.3.7. *Let (X, d_X) and (Y, d_Y) be compact and complete metric spaces respectively. Then the following two results are true.*

- (i) $C(X, Y)$ is a complete metric space with respect to the distance $d_{C(X,Y)}(\cdot, \cdot)$.
- (ii) A subset E of $C(X, Y)$ is compact if and only if E is bounded, closed, and equi-continuous in the sense that for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$d_X(x_1, x_2) < \delta \Rightarrow d_Y(f(x_1) - f(x_2)) < \varepsilon \text{ for all } f \in E.$$

Proof. (i) Suppose $\{f_k\}_{k=1}^\infty \subseteq C(X, Y)$ is a Cauchy sequence. Then for any $\varepsilon > 0$ there is an $n_0 \in \mathbf{N}$ such that if $k, l > n_0$ then $d_{C(X,Y)}(f_k, f_l) < \varepsilon$; that is, $d_Y(f_k(x), f_l(x)) < \varepsilon$ for all $x \in X$. Since Y is complete, Theorem 7.3.4 is applied to derive that $\{f_k\}_{k=1}^\infty$ converges uniformly on X to some mapping $f : X \rightarrow Y$. However, Theorem 7.3.5 indicates that f is continuous. Thus $f \in C(X, Y)$ and $\lim_{k \rightarrow \infty} f_k = f$ in the sense of points of the metric space $(C(X, Y), d_{C(X,Y)})$. Thus $C(X, Y)$ is complete under $d_{C(X,Y)}(\cdot, \cdot)$.

(ii) Let E be a subset of $C(X, Y)$. If E is compact, then E is automatically bounded and closed. Moreover, E is totally bounded by Theorem 7.3.7 and Corollary 6.3.6, and hence for $\varepsilon > 0$ there are finite many open balls $\{B_{3^{-1}\varepsilon}^{C(X,Y)}(f_k)\}_{k=1}^n$ of $C(X, Y)$ such that $E \subseteq \bigcup_{j=1}^n B_{3^{-1}\varepsilon}^{C(X,Y)}(f_k)$. Since each $f_k : X \rightarrow Y$ is uniformly continuous, there exists a $\delta_k > 0$ such that

$$d_X(x_1, x_2) < \delta_k \Rightarrow d_Y(f_k(x_1), f_k(x_2)) < 3^{-1}\varepsilon.$$

Let $\delta = \min\{\delta_k : k = 1, 2, \dots, n\}$. Then for a given $f \in E$, choose an f_k such that $d_{C(X,Y)}(f, f_k) < 3^{-1}\varepsilon$. Hence, $d_X(x_1, x_2) < \delta$ yields

$$d_Y(f(x_1), f(x_2)) \leq d_Y(f(x_1), f_k(x_1)) + d_Y(f_k(x_1), f_k(x_2)) + d_Y(f_k(x_2), f(x_2)) < \varepsilon.$$

Conversely, assume the statement after the if and only if. Since E is closed and $C(X, Y)$ is complete, according to Corollary 6.3.6 and Theorem 6.3.5 it remains to check that each sequence $\{f_k\}_{k=1}^\infty$ in E has a Cauchy subsequence. Since X is separable, there is a sequence $\{x_k\}_{k=1}^\infty$ which is dense in X . Note that E is bounded and then $\{f_k(x_1)\}_{k=1}^\infty$ is bounded in the complete space Y . So this sequence has a convergent subsequence, denoted $\{f_k^{(1)}(x_1)\}_{k=1}^\infty$. Similarly, suppose that $\{f_k^{(2)}\}_{k=1}^\infty$ is a subsequence of $\{f_k^{(1)}\}_{k=1}^\infty$ such that $\{f_k^{(2)}(x_2)\}_{k=1}^\infty$ converges. In this way, we can inductively obtain subsequence $\{f_k^{(j)}\}_{k=1}^\infty$ of $\{f_k\}_{k=1}^\infty$ such that $\lim_{k \rightarrow \infty} f_k^{(j)}(x_l)$ exists for $l = 1, 2, \dots, j$. Now $\{f_k^{(k)}\}_{k=1}^\infty$ ensures that $\{f_k^{(k)}(x_j)\}_{k=1}^\infty$ is convergent for each j . Since $f_k^{(k)} \in C(X, Y)$, for $\varepsilon > 0$ there is a $\delta > 0$ such that

$$d_X(x, y) < \delta \Rightarrow d_Y(f_k^{(k)}(x), f_k^{(k)}(y)) < 3^{-1}\varepsilon.$$

Also, note that $\{x_k\}_{k=1}^\infty$ is dense in the compact space X . So, it follows that $X = \bigcup_{l=1}^N B_\delta^X(x_l)$ for some natural number n_0 , where $B_\delta^X(x_l)$ is the open ball in X centered at x_l with radius δ . Furthermore, the definition of $f_k^{(k)}$ yields a natural number m_0 such that

$$m, n > m_0 \Rightarrow d_Y(f_m^{(m)}(x_j), f_n^{(n)}(x_j)) < 3^{-1}\varepsilon \text{ for } j = 1, 2, \dots, n_0.$$

Finally, upon choosing x_j such that $x \in B_\delta^X(x_j)$ for any $x \in X$, we get that if $m, n > m_0$ then

$$\begin{aligned} & d_Y(f_m^{(m)}(x), f_n^{(n)}(x)) \\ & \leq d_Y(f_m^{(m)}(x), f_m^{(m)}(x_j)) + d_Y(f_m^{(m)}(x_j), f_n^{(n)}(x_j)) + d_Y(f_n^{(n)}(x_j), f_n^{(n)}(x)) \\ & < \varepsilon, \end{aligned}$$

and consequently, $\{f_k^{(k)}\}_{k=1}^\infty$ is convergent to an element of E and hence E is compact. \square

Remark 7.3.8. *Theorem 7.3.7 extends (from a closed interval to a compact metric space) Guido Ascoli's theorem – every bounded equi-continuous sequence of real-valued functions on the unit interval $[0, 1]$ has a uniformly convergent subsequence.*

7.4 Contractions

In general, a fixed point of a mapping f from a set X to itself is a point $x_0 \in X$ such that $f(x_0) = x_0$. For example, 0 is a fixed point of the mapping $f(x) = x^2 + x$ from \mathbf{R} to \mathbf{R} . But, not all mappings have fixed points – for instance, the mapping $f(x) = x + 1$ has no fixed point on \mathbf{R} . So, what mapping has a fixed point is an interesting question. In what follows, we introduce an important tool – the so-called Banach fixed point theorem (named after Stefan Banach) – which ensures the existence and uniqueness of fixed points of certain self mappings of metric spaces, but also provides a constructive method to find those fixed points.

Definition 7.4.1. *Let (X, d_X) be a metric space. A mapping f from X to itself is a contraction if there is an $\alpha \in (0, 1)$ such that*

$$d_X(f(x_1), f(x_2)) \leq \alpha d_X(x_1, x_2) \text{ for all } x_1, x_2 \in X.$$

A contraction mapping contracts or shrinks the distance between points by the factor α . Clearly, any contraction map is uniformly continuous on X . This property actually suggests the following Banach's theorem on contraction mappings and fixed points.

Theorem 7.4.2. *Let (X, d_X) be a complete metric space. If $f : X \rightarrow X$ is a contraction mapping, then f has a unique fixed point.*

Proof. Let $x_0 \in X$. Define $x_{k+1} = f(x_k)$ for $k \in \mathbf{N} \cup \{0\}$. We claim that $\{x_k\}_{k=1}^\infty$ is a Cauchy sequence in X . In fact

$$d_X(x_2, x_1) = d_X(f(x_1), f(x_0)) \leq \alpha d_X(x_1, x_0),$$

and so

$$d_X(x_3, x_2) = d_X(f(x_2), f(x_1)) \leq \alpha d_X(x_2, x_1) \leq \alpha^2 d_X(x_1, x_0).$$

Generally, one has that if $k > j$ then

$$d_X(x_k, x_j) \leq \sum_{i=j}^{k-1} \alpha^i d_X(x_1, x_0) \leq \alpha^j (1 - \alpha)^{-1} d_X(x_1, x_0).$$

This, together with $\alpha \in (0, 1)$, implies $\{x_k\}_{k=1}^\infty$ is a Cauchy sequence and hence it converges to a point $x \in X$: $\lim_{k \rightarrow \infty} x_k = x$ in X . Since f is uniformly continuous, one has

$$f(x) = \lim_{k \rightarrow \infty} f(x_k) = \lim_{k \rightarrow \infty} x_{k+1} = x;$$

that is to say, x is a fixed point.

Regarding the uniqueness, suppose that y is also a fixed point of f . Then

$$d_X(x, y) = d_X(f(x), f(y)) \leq \alpha d_X(x, y)$$

and hence $d_X(x, y) = 0$; that is, $x = y$, thanks to $\alpha \in (0, 1)$. □

Remark 7.4.3. *The above proof is constructive in the sense that the fixed point is the limit of the iterates given by $x_{k+1} = f(x_k)$, where the initial point x_0 is an arbitrary point in X . The previous estimate gives the rapidity of the convergence $x_k \rightarrow x$:*

$$d_X(x, x_j) \leq \alpha^j (1 - \alpha)^{-1} d_X(f(x_0), x_0).$$

Below is an interesting example which shows an application of Theorem 7.4.2 in integral equations.

Example 7.4.4. *For a finite interval $[a, b]$ of \mathbf{R} , let $k : [a, b] \times [a, b] \rightarrow \mathbf{R}$ be continuous and $\phi \in C[a, b]$. An equation of the type:*

$$f(x) = \lambda \int_a^b k(x, y) f(y) dy + \phi(x), \quad \lambda \in \mathbf{R},$$

where f is an unknown function, is called a Fredholm integral equation of the second kind. This equation introduced by Ivar Fredholm has a unique solution $f \in C[a, b]$ for certain λ .

To check this example, define a mapping $T : C[a, b] \rightarrow C[a, b]$ by

$$T(f)(x) = \lambda \int_a^b k(x, y) f(y) dy + \phi(x) \quad \text{for } \lambda \in \mathbf{R}.$$

We write the distance function on $C[a, b]$ as

$$d_{C[a,b]}(f_1, f_2) = \max_{x \in [a,b]} |f_1(x) - f_2(x)| \quad \text{for all } f_1, f_2 \in C[a, b].$$

Solving the desired equation is equivalent to showing that T has a fixed point. Now

$$d_{C[a,b]}(T(f_1), T(f_2)) \leq |\lambda| \max_{x,y \in [a,b]} |k(x,y)|(b-a) d_{C[a,b]}(f_1, f_2).$$

Thus, if

$$|\lambda| \max_{x,y \in [a,b]} |k(x,y)|(b-a) < 1,$$

then T is a contraction and, therefore, has a unique fixed point by Theorem 7.4.2.

As a consequence of Theorem 7.4.2, we can get a more general fixed point theorem.

Corollary 7.4.5. *Let (X, d_X) be a complete metric space and $n - 1 \in \mathbf{N}$. If the n -th iteration $f_{(n)}(x) = \underbrace{f(f(f(\cdots(f(x))))}_{n}$ is a contraction, then f has a unique fixed point.*

Proof. Since $f_{(n)}$ is a contraction mapping, it has a unique fixed point x by Theorem 7.4.2. If α is the contraction constant for $f_{(n)}$, then

$$d_X(f(x), x) = d_X(f(f_{(n)}(x)), f_{(n)}(x)) = d_X(f_{(n)}(f(x)), f_{(n)}(x)) \leq \alpha d_X(f(x), x)$$

and hence $f(x) = x$.

If y is another fixed point of f then it is also a fixed point of $f_{(n)}$ since

$$f_{(n)}(y) = f_{(n-1)}(f(y)) = f_{(n-1)}(y) = \cdots = y.$$

The uniqueness of the fixed point of $f_{(n)}$ implies $y = x$. □

Remark 7.4.6. *In Corollary 7.4.5, f needs not be a contraction. In fact, there is a discontinuous function f from $[0, 1]$ to itself such that its second iteration becomes a contraction. For instance,*

$$f(x) = \begin{cases} 2^{-2} & , \quad x \in [0, 2^{-1}] \\ 2^{-1} & , \quad x \in (2^{-1}, 1] \end{cases} \Rightarrow f_{(2)}(x) = 2^{-2} \quad \text{for all } x \in [0, 1].$$

Example 7.4.7. Consider the following Volterra integral equation (VIE), named after Vito Volterra:

$$f(x) = \lambda \int_a^x k(x, y) f(y) dy + \phi(x),$$

for which the notations are the same as in Example 7.4.4. If $T : C[a, b] \rightarrow C[a, b]$ is defined by

$$T(f)(x) = \lambda \int_a^x k(x, y) f(y) dy + \phi(x),$$

then the VIE has a unique solution in $C[a, b]$ for any $\lambda \in \mathbf{R}$.

To see this assertion, it is enough to show that T in the above has a unique fixed point in $C[a, b]$. A simple computation implies that if

$$f_1, f_2 \in C[a, b] \quad \text{and} \quad M = \max_{x, y \in [a, b]} |k(x, y)|,$$

then

$$\begin{cases} |T(f_1)(x) - T(f_2)(x)| \leq |\lambda| M d_{C[a, b]}(f_1, f_2)(x - a); \\ |T_{(2)}(f_1)(x) - T_{(2)}(f_2)(x)| \leq 2^{-1} (|\lambda| M)^2 d_{C[a, b]}(f_1, f_2)(x - a)^2; \\ |T_{(n)}(f_1)(x) - T_{(n)}(f_2)(x)| \leq (n!)^{-1} (|\lambda| M)^n d_{C[a, b]}(f_1, f_2)(x - a)^n, \end{cases}$$

and hence

$$d_{C[a, b]}(T_{(n)}(f_1), T_{(n)}(f_2)) \leq (n!)^{-1} (|\lambda| (b - a) M)^n d_{C[a, b]}(f_1, f_2).$$

Since

$$\lim_{n \rightarrow \infty} (n!)^{-1} (|\lambda| (b - a) M)^n = 0,$$

$T_{(n)}$ is a contraction mapping for large n and, therefore, has a unique fixed point for any value of the parameter λ by Corollary 7.4.5.

As a consequence of Remark 7.4.3, we consider a family of contractions that vary continuously with respect to a parameter and prove that the corresponding fixed points also vary continuously.

Corollary 7.4.8. Let (X, d_X) be a complete metric space and let (Y, d_Y) be a metric space. Suppose that $f : Y \times X \rightarrow X$ is such that $f(\cdot, x)$ is continuous on Y for any point $x \in X$ and there exists $\alpha \in (0, 1)$ such that

$$d_X(f(y, x_1), f(y, x_2)) \leq \alpha d_X(x_1, x_2) \quad \text{for all } y \in Y \text{ and } x_1, x_2 \in X.$$

For $y \in Y$ let $x_y \in X$ be the unique fixed point of the contraction $f(y, \cdot)$. Then the mapping $y \mapsto x_y$ is continuous from Y to X .

Proof. For $y \in Y$, consider the equation $x = f(y, x)$. Let $y_0 \in Y$. Construct a solution of the equation starting with the initial point $x_0 = x_{y_0}$ so that $x_{k+1} = f(y, x_k) \rightarrow x_y$. From Remark 7.4.3 it follows that

$$d_X(x_{y_0}, x_y) \leq (1 - \alpha)^{-1} d_X(x_{y_0}, f(y, x_{y_0})) = (1 - \alpha)^{-1} d_X(f(y_0, x_{y_0}), f(y, x_{y_0}))$$

so that the mapping $y \rightarrow x_y$ is continuous at y_0 since the mapping $y \mapsto f(y, x_{y_0})$ is continuous at y_0 . \square

As a further application of the contraction mapping theorem, we give the following Charles Emile Picard's existence and uniqueness theorem for a first-order system of nonlinear ordinary differential equations based on the Rudolf Lipschitz continuity.

Theorem 7.4.9. For $n \in \mathbf{N}$ equip \mathbf{R}^n with the distance

$$d_n(x, y) = \|x - y\|_n = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$$

between both $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in \mathbf{R}^n . Let $O \subseteq \mathbf{R}^n$ be open and $[a, b] \subseteq \mathbf{R}$ be a finite interval. Suppose $f : [a, b] \times O \rightarrow \mathbf{R}^n$ is a continuous and satisfies a Lipschitz condition with the constant $\kappa > 0$ below:

$$d_n(f(t, y_1), f(t, y_2)) \leq \kappa d_n(y_1, y_2) \text{ for all } (t, y_1), (t, y_2) \in [a, b] \times O.$$

Then, there exists a $\delta \in (a, b]$ such that the initial value problem (IVP):

$$\frac{d}{dt}u(t) = u'(t) = f(t, u(t)) \text{ subject to } u(a) = y_0 \in \mathbf{R}^n$$

has a unique solution in the interval $[a, \delta]$.

Proof. Clearly, solving this IVP is equivalent to solving the integral equation

$$u(t) = y_0 + \int_a^t f(s, u(s)) ds.$$

So, it is our aim to show the existence and uniqueness of a solution to this integral equation.

Assuming that $\overline{B}_r(y_0) \subseteq O$ is the closed ball centered at y_0 with radius $r > 0$, we have

$$M = \max_{(t, y) \in [a, b] \times \overline{B}_r(y_0)} \|f(t, y)\|_n < \infty.$$

Now choose $\delta \in (a, b]$ such that $(\delta - a)\kappa < 1$ and $(\delta - a)M \leq r$. Set S be the class of continuous mappings from $[a, \delta]$ to $\overline{B}_r(y_0)$. It is clear that S is complete under the distance defined by

$$d_S(\phi_1, \phi_2) = \max_{t \in [a, \delta]} d_n(\phi_1(t), \phi_2(t)) \text{ for all } \phi_1, \phi_2 \in S.$$

Define a mapping $F : S \rightarrow S$ by

$$F(\phi)(t) = y_0 + \int_a^t f(s, \phi(s)) ds.$$

Obviously, $F(\phi)$ is continuously differentiable. If $\phi \in S$ and $t \in [a, \delta]$, then

$$d_n(F(\phi)(t), y_0) = \left\| \int_a^t f(s, \phi(s)) ds \right\|_n \leq M|t - a| \leq (\delta - a)M \leq r.$$

This implies $F(\phi) \in S$. Now, our problem amounts to showing that F has a unique fixed point in S . Thus, it suffices to verify that F is a contraction. To do so, we note that $t \in [a, \delta]$ and $\phi_1, \phi_2 \in S$ imply

$$d_n(F(\phi_1)(t), F(\phi_2)(t)) \leq \int_a^t \|f(s, \phi_1(s)) - f(s, \phi_2(s))\|_n ds \leq (\delta - a)\kappa d_S(\phi_1, \phi_2),$$

whence yielding

$$d_S(F(\phi_1), F(\phi_2)) \leq (\delta - a)\kappa d_S(\phi_1, \phi_2).$$

Accordingly, the assertion follows. □

Remark 7.4.10.

(i) *Note that any existence theorem for the nonlinear IVP in the above must be local in nature. For example,*

$$y'(t) = y^2(t) \text{ subject to } y(1) = -1$$

has the solution $y(t) = -1/t$, which is not defined at $t = 0$ even though $f(t, y) = y^2$ is satisfied with the required condition on $O \subseteq \mathbf{R}$ which is assumed to be open and bounded.

(ii) If the Lipschitz condition is dropped, then it is still possible for the IVP to have one and even many more solutions – for instance, the IVP:

$$y'(t) = y^{\frac{1}{3}}(t) \text{ subject to } y(0) = 0$$

has an infinite number of solutions

$$y_c(t) = \begin{cases} 0 & , 0 \leq t \leq c \\ \left(\left(\frac{2}{3}\right)(t-c)\right)^{\frac{3}{2}} & , c < t \leq 1 \end{cases} \quad \text{where } c \in [0, 1].$$

7.5 Structures of Metric Spaces

It is often necessary to compare two metric spaces and decide in what sense they are equivalent or to analyze how the structure of a metric space changes when changing the metric. Note that every metric space is a set with additional topological structure induced by the metric. So to decide in what sense two metric spaces are equivalent we have to use continuous mappings between them.

Definition 7.5.1. Let (X, d_X) and (Y, d_Y) be two metric spaces. Then they are called:

- (i) *topologically isomorphic or homeomorphic* provided that there exists a homeomorphism between them – a mapping $f : X \rightarrow Y$ which is bijective and continuous, and has continuous inverse;
- (ii) *uniformly isomorphic* provided that there exists a uniform isomorphism between them – a mapping $f : X \rightarrow Y$ which is bijective and uniformly continuous, and has uniformly continuous inverse;
- (iii) *isometrically isomorphic* provided that there exists a bijective mapping f such that $f : X \rightarrow Y$ and $d_Y(f(x_1), f(x_2)) = d_X(x_1, x_2)$ for all $x_1, x_2 \in X$;
- (iv) *similar* provided that there exists a positive constant $\kappa > 0$ and a bijective mapping f such that $f : X \rightarrow Y$ and $d_Y(f(x_1), f(x_2)) = \kappa d_X(x_1, x_2)$ for all $x_1, x_2 \in X$.
- (v) *equivalent* provided that as sets $X = Y$ and there exist two positive constants κ_1 and κ_2 independent of all $x_1, x_2 \in X$ such that

$$\kappa_1 d_X(x_1, x_2) \leq d_Y(x_1, x_2) \leq \kappa_2 d_X(x_1, x_2).$$

The following example helps us grasp Definition 7.5.1.

Example 7.5.2.

(i) Let $f : [-2^{-1}\pi, 2^{-1}\pi] \rightarrow [-1, 1]$ be defined by $f(x) = \sin x$. Then it is a uniform isomorphism, and hence any two finite closed subintervals of \mathbf{R} are uniformly isomorphic.

(ii) For $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbf{R}^n$ and $n \in \mathbf{N}$, recall

$$d_n(x, y) = \|x - y\|_n = \sqrt{\sum_{k=1}^n |x_k - y_k|^2}$$

and let

$$d_\infty(x, y) = \|x - y\|_\infty = \max\{|x_k - y_k| : k = 1, \dots, n\}.$$

Then

$$d_n(x, y) \leq d_\infty(x, y) \leq n d_n(x, y).$$

This means that (\mathbf{R}^n, d_n) and (\mathbf{R}^n, d_∞) are equivalent. Of course, d_n and d_∞ define the same notions of continuity and convergence and do not need to be distinguished for most purposes. In other words, the identity mapping is uniformly isomorphic from (\mathbf{R}^n, d_n) to (\mathbf{R}^n, d_∞) .

(iii) In (\mathbf{R}^n, d_n) , an isometrically isomorphic mapping is a rotation (the movement of a body in such a way that the distance between a certain fixed point and any given point of that body remains constant), a reflection (to invert a geometric figure, respect to a line or plane, but not a point), or a translation (to move every point by a fixed distance in the same direction).

(iv) (\mathbf{R}^n, d_n) is similar to itself. This is because the mapping $f(x) = \kappa x + x_0$, where $\kappa > 0$ is constant and $x_0 \in \mathbf{R}^n$, is bijective and satisfies $d_n(f(x), f(y)) = \kappa d_n(x, y)$ for any $x, y \in \mathbf{R}^n$.

Both a uniform isomorphism and an isometric isomorphism are certainly homeomorphisms which are not only open mappings (from open sets to open sets) but also closed mappings (from closed sets to closed sets). Intuitively, a homeomorphism not only maps points in the first object that are close together to points in the second object that are close together, but also sends points in the first object that are not close together to points in the second object that are not close together. Topology is the study of those properties of objects that do not change when homeomorphisms are applied. Below is a classical and important result.

Theorem 7.5.3. For $n+1 \in \mathbf{N}$ let \mathbf{R}^{n+1} be equipped with the Euclidean distance

$$d_{n+1}(x, y) = \|x - y\|_{n+1} = \sqrt{\sum_{k=1}^{n+1} (x_k - y_k)^2}$$

for $x = (x_1, \dots, x_{n+1})$ and $y = (y_1, \dots, y_{n+1})$ in \mathbf{R}^{n+1} .

(i) If $n = 0$, then \mathbf{R}^{n+1} is homeomorphic to the open interval $(-1, 1) \subseteq \mathbf{R}$.

(ii) If $n \geq 1$, then \mathbf{R}^n is homeomorphic to the punctured sphere $\mathbf{S}^n \setminus \{(0, \dots, 1)\}$, where

$$\mathbf{S}^n = \{x \in \mathbf{R}^{n+1} : \|x\|_{n+1} = d_{n+1}(x, 0) = 1\}$$

is the compact unit sphere in \mathbf{R}^{n+1} .

Proof. To check (i), we just take $f(x) = 2\pi^{-1} \arctan x$.

As with (ii), we define the mapping $f : \mathbf{S}^n \setminus \{(0, \dots, 1)\} \rightarrow \mathbf{R}^n$ by letting

$$y = (y_1, \dots, y_n) = f((x_1, \dots, x_{n+1})) = f(x)$$

be the point in \mathbf{R}^n such that

$$y_k = x_k(1 - x_{n+1})^{-1} \text{ for } k = 1, 2, \dots, n.$$

Next we define $h : \mathbf{R}^n \rightarrow \mathbf{R}^{n+1}$ by letting $x = h(y)$ be the point in \mathbf{R}^{n+1} such that

$$x_k = \begin{cases} \frac{2y_k}{1+d_n^2(y, 0)}, & k = 1, \dots, n \\ \frac{d_n^2(y, 0)-1}{d_n^2(y, 0)+1}, & k = n+1 \end{cases}.$$

Note that $d_{n+1}(h(y), 0) = 1$ and $x_{n+1} \neq 1$. So, $h(\mathbf{R}^n) \subseteq \mathbf{S}^n \setminus \{(0, 0, \dots, 1)\}$. It is easy to see that both f and h are continuous and satisfy $f \circ h(y) = y$ and $h \circ f(x) = x$. Thus, h is the inverse of f , and f is a homeomorphism of $\mathbf{S}^n \setminus \{(0, \dots, 1)\}$ onto \mathbf{R}^n .

Regarding the compactness of \mathbf{S}^n , we define a function $F : \mathbf{R}^{n+1} \rightarrow \mathbf{R}$ via $F(x) = \|x\|_{n+1}$. The triangle inequality for d_{n+1} yields that F is continuous. With this and the closedness of the single point $\{1\}$, we see that $\mathbf{S}^n = F^{-1}(\{1\})$ is closed. Since $d_{n+1}(0, x) = 1$ for $x \in \mathbf{S}^n$, \mathbf{S}^n is bounded. Being closed, \mathbf{S}^n is therefore compact. \square

Problems

7.1. Let $f : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ be given by $f(x) = (x_1, x_2)$ for $x = (x_1, x_2, x_3) \in \mathbf{R}^3$. Prove that f is continuous on \mathbf{R}^3 .

7.2. Given $f(x) = d_X(x, 0)$ for every point x in the subset $X \setminus \{0\}$ of a metric space X with distance d_X , define $f(0)$ so that f is continuous at 0.

7.3. For $n \in \mathbf{N}$ let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be given by $f(x) = (\sum_{j=1}^n x_j^2)^{2^{-1}}$ for $x = (x_1, \dots, x_n) \in \mathbf{R}^n$, and let $S = (-3, -2)$. Find $f^{-1}(S)$.

7.4. Define $D = \{x = (x_1, x_2, x_3) \in \mathbf{R}^3 : 0 < x_3 < x_1^2 + x_2^2\}$ and let f be the mapping from \mathbf{R}^3 into \mathbf{R} given by

$$f(x) = \begin{cases} 1, & x \in D \\ 0, & x \notin D \end{cases}.$$

Show that $f(x^{(k)})$ tends to $f(0)$ as $x^{(k)}$ approaches 0 along any linear path, but f is discontinuous at 0.

7.5. Let T be the linear transformation from \mathbf{R}^2 into itself such that

$$T((1, 1)) = (3, -1) \text{ and } T((1, -1)) = (1, 7).$$

Find the matrix representation of T .

7.6. Given $n \in \mathbf{N}$. If K is a compact subset of \mathbf{R}^n , $f : K \rightarrow \mathbf{R}$ is continuous, and $f(x) > 0$ for any $x \in K$, show that there is a $\delta > 0$ such that $f(x) \geq \delta$ for any $x \in K$.

7.7. Show that the function $f : [0, 1] \rightarrow [0, 1]$, defined by

$$f(x) = \begin{cases} x & , x \in \mathbf{Q} \cap [0, 1] \\ 1 - x & , x \in (\mathbf{R} \setminus \mathbf{Q}) \cap [0, 1] \end{cases}$$

is onto and thus satisfies the intermediate value property. Show, however, that f is continuous only at $x = 2^{-1}$.

7.8. Define

$$T : C[0, 1] \rightarrow C[0, 1] \text{ by } T(f)(x) = \int_0^x f(t) dt.$$

Verify the following three facts:

- (i) T is not a contraction;
- (ii) T has a unique fixed point;
- (iii) $T \circ T$ is a contraction.

7.9. Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be differentiable with $|f'(x)| \leq \alpha$, where $\alpha \in (0, 1)$. Prove that f is a contraction mapping.

7.10. For a complete metric space (X, d_X) let $T : X \rightarrow X$ be a mapping and

$$\sum_{k=1}^{\infty} \sup_{x_1 \neq x_2} \frac{d_X(T_{(k)}(x_1), T_{(k)}(x_2))}{d_X(x_1, x_2)} < \infty.$$

Prove that T has a unique fixed point, where $T_{(k)} = \underbrace{T \circ T \cdots \circ T}_k$.

7.11. Let (X, d_X) be a metric space. Suppose K is a nonempty compact subset of X and the mapping $T : K \rightarrow K$ satisfies

$$d_X(T(x_1), T(x_2)) < d_X(x_1, x_2) \text{ for all } x_1 \neq x_2 \text{ in } K,$$

prove that there exists precisely one point $x \in K$ such that $x = T(x)$.

7.12. Let $h \in C[0, 1]$. Show that there is an $f \in C[0, 1]$ such that

$$f(x) - \int_0^x f(x-t) \exp(-t^2) dt = h(x).$$

7.13. Let (X, d_X) be a complete metric space and $f : D \subseteq X \rightarrow X = f(D)$ be a mapping. Suppose that there is a number $\kappa > 1$ such that

$$d_X(f(x_1), f(x_2)) \geq \kappa d_X(x_1, x_2) \text{ for all } x_1, x_2 \in D.$$

Prove that there exists exactly one point $x \in D$ such that $f(x) = x$.

7.14. Let (X, d_X) and (Y, d_Y) be two metric spaces and $f : X \rightarrow Y$ bijective. Prove that the following three conditions are equivalent:

- (i) f is a homeomorphism;
- (ii) f is continuous, and maps closed subsets of X to closed subsets of Y ;
- (iii) f is continuous and maps open subsets of X to open subsets of Y .

7.15. Suppose that \mathbf{R} is equipped with the usual Euclidean distance. Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be defined by

$$f(x) = \begin{cases} x + 3 \cdot 2^{-1}, & x \in (-\infty, 2^{-1}] \\ x^{-1}, & x \in (2^{-1}, \infty) \end{cases}.$$

Prove that f is not homeomorphic.

7.16. Let (X, d_X) be a metric space. Prove that there is a complete metric space (Y, d_Y) and a mapping $f : X \rightarrow Y$ such that f is an isometry, i.e.,

$$d_Y(f(x_1), f(x_2)) = d_X(x_1, x_2) \text{ for all } x_1, x_2 \in X,$$

and $f(X)$ is a dense subset of Y .

7.17. Let (X, d_X) be a metric space and K be a nonempty closed subset of X . Prove that if $f : X \rightarrow \mathbf{R}$ is defined by $f(x) = \inf_{y \in K} d_X(x, y)$ then:

(i) f is continuous;

(ii) $f(x) = 0$ if and only if $x \in K$.

Chapter 8

Normed Linear Spaces

In contrast to the previous practice of examining functions individually, the theory of functionals is initially treated as studying common and distinct properties of various spaces of functions, but actually it can be considered as proceeding from an interplay between both the linear algebra and the Euclidean topology, resulting in the notion of normed linear space. Accordingly we begin our discussion with this fundamental concept which, when combined with increasingly specialized assumptions, will lead to various increasingly specialized spaces and results. Here we would like to say that our investigation will involve only linear functional analysis whose relation to the developing nonlinear counterpart is similar to and arises from the connection between the Euclidean spaces and manifolds.

8.1 Linear Spaces, Norms and Quotient Spaces

From now on, we always use \mathbf{F} as either \mathbf{R} or $\mathbf{C} = \mathbf{R} + i\mathbf{R}$ – the field of real or complex numbers.

Definition 8.1.1. *A linear (or vector) space over \mathbf{F} is a set X equipped with two functions*

$$+ : X \times X \rightarrow X \text{ and } \cdot : \mathbf{F} \times X \rightarrow X$$

with the properties below:

- (i) $x + y = y + x$ for all $x, y \in X$;
- (ii) $(x + y) + z = x + (y + z)$ for all $x, y, z \in X$;

- (iii) *there is $0 \in X$ such that $x + 0 = 0 + x = x$ for all $x \in X$;*
- (iv) *there is $-x \in X$ such that $x + (-x) = (-x) + x = 0$ for all $x \in X$;*
- (v) $\alpha \cdot (\beta \cdot x) = (\alpha\beta) \cdot x$ *for all $\alpha, \beta \in \mathbf{F}, x \in X$;*
- (vi) $(\alpha + \beta) \cdot x = (\alpha \cdot x) + (\beta \cdot x)$ *for all $\alpha, \beta \in \mathbf{F}, x \in X$;*
- (vii) $\alpha \cdot (x + y) = (\alpha \cdot x) + (\alpha \cdot y)$ *for all $\alpha \in \mathbf{F}, x, y \in X$;*
- (viii) $1 \cdot x = x$ *for all $x \in X$ where 1 is the multiplicative identity in \mathbf{F} .*

Although our main purpose is to study linear spaces over \mathbf{R} or \mathbf{C} , we frequently state that the spaces are over \mathbf{F} whenever both \mathbf{R} and \mathbf{C} may be handled in the same way.

Example 8.1.2.

- (i) *For $n \in \mathbf{N}$, \mathbf{F}^n is a linear space with the usual vector addition and scalar multiplication over \mathbf{F} .*
- (ii) *Let X be the set of all polynomials with coefficients in \mathbf{R} of degree less than $n \in \mathbf{N}$. Then X is a linear space with usual addition of polynomials and scalar multiplication over \mathbf{R} .*
- (iii) *For $m, n \in \mathbf{N}$ let $M_{m,n}(\mathbf{C})$ be the set of all complex-valued $m \times n$ matrices. Then $M_{m,n}(\mathbf{C})$ is a linear space with usual addition of matrices and scalar multiplication over \mathbf{C} .*
- (iv) *Recall that ℓ_∞ represents the set of infinite real-valued sequences $(x_j)_{j=1}^\infty$ which are bounded: $\sup_{j \in \mathbf{N}} |x_j| < \infty$. Then ℓ_∞ is a linear space over \mathbf{R} with:*

$$\sup_{j \in \mathbf{N}} |x_j + y_j| \leq \sup_{j \in \mathbf{N}} |x_j| + \sup_{j \in \mathbf{N}} |y_j|; \quad \sup_{j \in \mathbf{N}} |\alpha x_j| = |\alpha| \sup_{j \in \mathbf{N}} |x_j| < \infty.$$

- (v) *Let $C(S, \mathbf{R})$ be the class of all continuous functions $f : S \rightarrow \mathbf{R}$ with*

$$(f + g)(x) = f(x) + g(x) \quad \text{and} \quad (\alpha \cdot f)(x) = \alpha f(x).$$

Here S is any nonempty subset of \mathbf{R} . Then $C(S, \mathbf{R})$ is a linear space over \mathbf{R} .

- (vi) *Let $C^\infty[a, b]$ be the space of all infinitely differentiable real-valued functions on the finite interval $[a, b]$ of \mathbf{R} . Then it is linear space over \mathbf{R} with the same addition and scalar multiplication as in (v).*
- (vii) *For any closed interval $[a, b] \subset \mathbf{R}$ and any increasing function $g : \mathbf{R} \rightarrow \mathbf{R}$, the classes $R[a, b]$, $RS_g[a, b]$ and $LRS_g([a, b])$ are also linear spaces over \mathbf{R} with the same addition and scalar multiplication as in (v).*

For convenience, we will drop the special notation $+$, \cdot for vector addition and scalar multiplication, and simply refer to X as a linear space over \mathbf{F} . Moreover, if $\mathbf{F} = \mathbf{R}$ then we will say that X is a real linear space; whereas if $\mathbf{F} = \mathbf{C}$, then we will say that X is a complex linear space.

As in the linear algebra of finite-dimensional vector spaces, we have the concept of a linear subspace which is often called simply a subspace (when the context serves to distinguish it from other kinds of subspaces) and plays an important role in functional analysis and related fields of mathematics.

Definition 8.1.3. *Let X be a linear space over \mathbf{F} .*

(i) *A subset $Y \subseteq X$ is called a linear subspace of X provided that*

$$\alpha, \beta \in \mathbf{F} \text{ and } x, y \in Y \Rightarrow \alpha x + \beta y \in Y.$$

(ii) *If $Y = \{x + x_0 : x \in S\}$ where S is a linear subspace of X and x_0 is a fixed element of X , then Y is called an affine subset of X .*

Example 8.1.4.

(i) *For $n - 2 \in \mathbf{N}$, the vectors in \mathbf{R}^n of the form $(x_1, x_2, x_3, 0, \dots, 0)$ form a linear subspace of \mathbf{R}^n .*

(ii) *For $m, n \in \mathbf{N}$ and $m \leq n$, the set of polynomials of degree $\leq m$ forms a linear subspace of the set of polynomials of degree $\leq n$.*

(iii) *If $Y = \{(x_1, x_2, 1, \dots, 1) \in \mathbf{R}^n\}$ for $n - 2 \in \mathbf{N}$, then Y is an affine subset of \mathbf{R}^n .*

(iv) *In Example 8.1.2 (iii), let Y be the set of matrices with certain blocks of 1's. Then Y is an affine subset of $M_{m,n}(\mathbf{C})$.*

(v) *In \mathbf{R}^3 all lines and planes through the origin are linear subspaces, whereas all lines and planes not passing through the origin are affine subsets.*

A fundamental concept for linear spaces is that of dimension. To see this, suppose that X is a linear space over \mathbf{F} and $n \in \mathbf{N}$. Elements e_1, e_2, \dots, e_n of X are linearly dependent provided there are scalars $\alpha_1, \alpha_2, \dots, \alpha_n$ (not all zero) such that $\sum_{j=1}^n \alpha_j e_j = 0$. If there is no such set of scalars, then they are linearly independent. The linear span of the vectors $\{e_j\}_{j=1}^n$ is the linear subspace of X as follows:

$$\text{span}\{e_j\}_{j=1}^n = \text{span}\{e_1, \dots, e_n\} = \left\{ \sum_{j=1}^n \alpha_j e_j : \alpha_j \in \mathbf{F} \right\}.$$

In general, the linear span of a subset S of X is defined to be the set of all finite linear combinations of elements in S ; that is, the intersection of all linear subspaces of X containing S .

Definition 8.1.5. *Let $n \in \mathbf{N}$. If the linear space X is equal to the space spanned by a linearly independent set of n vectors in X , then X is said to have the dimension n . If there is no such set of vectors, then X is infinite-dimensional. Furthermore, a linearly independent set of vectors that spans X is called a basis for X .*

Example 8.1.6.

- (i) For $n \in \mathbf{N}$, the space \mathbf{R}^n has dimension n ; the standard basis is given by the vectors $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, ..., $e_n = (0, 0, \dots, 0, 1)$.
- (ii) For $n \in \mathbf{N}$, the set $\{1, t, t^2, \dots, t^n\}$ is a basis for the linear space of real-valued polynomials of degree $\leq n$ which has dimension $n + 1$.
- (iii) All linear spaces given in Example 8.1.2 (iv)-(vii) are infinite-dimensional.

Next, we equip a linear space with a norm. The norm on a linear space is a way of measuring the length of a vector and hence the distance between two vectors.

Definition 8.1.7. *Let X be a linear space over \mathbf{F} . Then a norm on X is a non-negative function $\|\cdot\| : X \rightarrow \mathbf{R}$ with the following three properties:*

- (i) $\|x\| = 0 \Leftrightarrow x = 0$;
- (ii) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$;
- (iii) $\|\alpha x\| = |\alpha| \|x\|$ for all $x \in X$ and $\alpha \in \mathbf{F}$.

In this case, $(X, \|\cdot\|)$ is called a normed linear space.

Remark 8.1.8.

- (i) In the definition we are assuming that $|\cdot|$ denotes the usual absolute value. If $\|\cdot\|$ is a nonnegative function only with both (ii) – the triangle inequality and (iii) – homogeneity, then it is called a semi-norm. For instance, $\|\cdot\|_{p, m_g, E}$ defined in Example 6.1.3 (iii) is a semi-norm whenever $p \in [1, \infty)$.
- (ii) Whenever $(X, \|\cdot\|)$ is a normed linear space over \mathbf{F} , defining

$$d_X(x_1, x_2) = \|x_1 - x_2\| \text{ for } x_1, x_2 \in X$$

we find that (X, d_X) is a metric space since d_X is a distance function which can be sufficiently demonstrated by

$$d_X(x_1, x_3) = \|x_1 - x_3\| \leq \|x_1 - x_2\| + \|x_2 - x_3\| = d_X(x_1, x_2) + d_X(x_2, x_3).$$

(iii) Clearly, if Y is a linear subspace of the linear space X (over \mathbf{F}) which is equipped with norm $\|\cdot\|$, then $(Y, \|\cdot\|)$ is also a normed linear space and hence it can be regarded as a linear subspace of $(X, \|\cdot\|)$.

Example 8.1.9.

(i) For $n \in \mathbf{N}$ and $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ let $d_n(x, 0) = \sqrt{\sum_{j=1}^n |x_j|^2}$. Then $d_n(\cdot, 0)$ defines a norm on \mathbf{R}^n – the only difficulty to verify this fact is the triangle inequality – for this we use the Cauchy-Schwarz inequality:

$$\sum_{j=1}^n |x_j y_j| \leq \sqrt{\left(\sum_{j=1}^n |x_j|^2\right) \left(\sum_{j=1}^n |y_j|^2\right)}.$$

(ii) For $n \in \mathbf{N}$, there are many other norms on \mathbf{R}^n , called the (n, p) -norms. More precisely, if $p \in [1, \infty)$, then

$$\|x\|_{n,p} = \left(\sum_{j=1}^n |x_j|^p\right)^{\frac{1}{p}}$$

is a norm on \mathbf{R}^n . To see this, it suffices to verify the triangle inequality, i.e., the following Hermann Minkowski's inequality (see also Problem 8.1 (i))

$$\|x + y\|_{n,p} \leq \|x\|_{n,p} + \|y\|_{n,p} \text{ for } x, y \in \mathbf{R}^n.$$

If $p = \infty$, then

$$\|x\|_{n,\infty} = \sup_{1 \leq j \leq n} |x_j|$$

defines a norm on \mathbf{R}^n . It is conventional to write ℓ_p^n for these normed spaces. Note that ℓ_p^n and ℓ_q^n have exactly the same elements which are just ones of \mathbf{R}^n .

(iii) Recall that ℓ_∞ is the linear space of bounded infinite sequences of real numbers. For $x = (x_j)_{j=1}^\infty \in \ell_\infty$, define

$$\|x\|_p = \begin{cases} \left(\sum_{j=1}^\infty |x_j|^p\right)^{\frac{1}{p}}, & 1 \leq p < \infty \\ \sup_{1 \leq j < \infty} |x_j|, & p = \infty \end{cases}.$$

If our attention is restricted to the linear subspace on which $\|\cdot\|_p$ is finite, then $\|\cdot\|_p$ is a norm – the somewhat difficult triangle inequality follows from the Minkowski inequality for the infinite sequences which can be established by the Minkowski inequality for \mathbf{R}^n in (ii) – see also Problem 8.1 (ii). This yields an infinite family of normed linear spaces:

$$\ell_p = \{x \in \ell_\infty : \|x\|_p < \infty\} \text{ for } p \in [1, \infty].$$

Let $p < q \in [1, \infty]$. Then there is a strict inclusion: $\ell_p \subseteq \ell_q$ and hence ℓ_p is a linear subspace of ℓ_q . Moreover, $(\ell_p, \|\cdot\|_p)$ is separable for $p \in [1, \infty)$ – for a proof see also Example 6.3.11 (ii).

(iv) For a finite closed interval $[a, b]$ in \mathbf{R} , $C[a, b]$ becomes a normed space with the p -norm below

$$\|f\|_p = \begin{cases} \left(\int_a^b |f(t)|^p dt \right)^{\frac{1}{p}}, & 1 \leq p < \infty \\ \sup_{a \leq t \leq b} |f(t)|, & p = \infty \end{cases}.$$

Note that the triangle inequality for $p \in [1, \infty)$ follows from the Minkowski inequality $\|f_1 + f_2\|_p \leq \|f_1\|_p + \|f_2\|_p$ – see also Remark 4.5.7.

As an application of the above concepts, quotients of normed linear spaces may be formulated. Note that we need both the algebraic structure (linear subspace of a linear space) and a topological property (closedness) to make it all work.

Definition 8.1.10. Let X be a linear space over \mathbf{F} and Y be a linear subspace of X . Then the linear space X/Y – the quotient or factor space is formed as follows. The elements of X/Y are cosets of Y – sets of the form

$$x + Y = \{x + y : y \in Y\} \text{ where } x \in X.$$

The set of cosets is a linear space under the operations:

$$(x_1 + Y) + (x_2 + Y) = x_1 + x_2 + Y \text{ and } \lambda \cdot (x + Y) = \lambda x + Y$$

for any $x_1, x_2, x \in X$ and $\lambda \in \mathbf{F}$. In addition, $\dim(X/Y)$ is called the codimension of a linear subspace $Y \subseteq X$.

Remark 8.1.11. *The above-defined two operations make sense precisely because Y is itself a linear space, so for instance,*

$$Y + Y = \{y_1 + y_2 : y_1, y_2 \in Y\} = Y \text{ and } \lambda Y = \{\lambda y : y \in Y\} = Y \text{ for } \lambda \neq 0.$$

Moreover, $x_1 + Y$ and $x_2 + Y$ are equal if and only if as sets $x_1 + Y = x_2 + Y$; this is true if and only if $x_1 - x_2 \in Y$, denoted $x_1 \sim_Y x_2$, which reads: x_1 is equivalent to x_2 with respect to Y .

Example 8.1.12.

(i) Let $X = \mathbf{R}^3$, and let Y be the linear subspace spanned by $(1, 1, 0)$. Then X/Y is a two-dimensional real vector space since $(1, 0, 1) + Y$ and $(0, 0, 1) + Y$ generate X/Y .

(ii) The linear space Y of finitely supported sequences (of which all but a finite number of entries vanish) in ℓ_1 is a linear subspace. The quotient space ℓ_1/Y is quite hard to visualize – its elements are equivalence classes under the relation $(x_j)_{j=1}^\infty \sim_Y (y_j)_{j=1}^\infty$ whenever $(x_j)_{j=1}^\infty$ and $(y_j)_{j=1}^\infty$ differ in finitely many coordinates.

(iii) For $n \in \mathbf{N}$ let Y be the set of ℓ_1 -sequences of $(0, \dots, 0, x_{n+1}, \dots)$ (first n coordinates are zero). Then Y is a linear subspace of ℓ_1 . Here the quotient space ℓ_1/Y is isomorphic to \mathbf{R}^n .

(iv) Recall that if $p, q \in [1, \infty]$ and $p < q$ then $\ell_p \subseteq \ell_q$ – this implies that ℓ_q/ℓ_p exists as a linear quotient space.

(v) Let $X = C[0, 1]$ and $Y = \{f \in X : f(0) = 0\}$. Then X/Y is isomorphic to \mathbf{R} .

(vi) $Y = C[0, 1]$ is a linear subspace of $X = R[0, 1]$. The quotient X/Y is again a linear space.

Evidently, these examples indicate that not all linear subspaces are equally good: Example 8.1.12 (i), (iii) and (v) are quite reasonable, whereas Example 8.1.12 (ii), (iv) and (vi) are examples of linear spaces unlike any we have seen. The reason is the following: The space X/Y is guaranteed to be a normed space with a norm related to the original norm $\|\cdot\|$ on X only when the subspace Y is itself closed. Notice that Example 8.1.12 (i), (iii) and (v) are precisely the ones in which the linear subspace is closed whenever the space X is treated as a metric space with the distance

$$d_X(x_1, x_2) = \|x_1 - x_2\| \text{ for all } x_1, x_2 \in X.$$

Theorem 8.1.13. *Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} and Y a linear subspace of X . Then X/Y is a normed linear space over \mathbf{F} under the norm*

$$\|x+Y\|_{X/Y} = \inf_{z \in x+Y} \|z\|_X$$

when and only when Y is a closed subset of X ; that is, if $\{y_j\}_{j=1}^\infty$ is a sequence in Y with $\lim_{j \rightarrow \infty} \|y_j - y\|_X = 0$ then $y \in Y$.

Proof. On the one hand, suppose that Y is closed and note that

$$\|x+Y\|_{X/Y} = \inf_{z \in x+Y} \|z\|_X = \inf_{y \in Y} \|x-y\|_X.$$

Firstly, if $\|x+Y\|_{X/Y} = 0$ then there is a sequence of vectors $y_j \in Y$ such that $\lim_{j \rightarrow \infty} \|x - y_j\|_X = 0$. Since Y is closed, $x \in Y$ and then $x+Y = 0+Y$. Consequently, $\|0+Y\|_{X/Y} = 0$.

Secondly, the homogeneity is clear from $\lambda \in \mathbf{F}$ and

$$\|\lambda(x+Y)\|_{X/Y} = \inf_{z \in x+Y} \|\lambda z\|_X = |\lambda| \inf_{z \in x+Y} \|z\|_X = |\lambda| \|x+Y\|_{X/Y}.$$

Finally, the triangle inequality is clear from

$$\begin{aligned} \|(x_1+Y) + (x_2+Y)\|_{X/Y} &= \inf_{z_1 \in x_1+Y, z_2 \in x_2+Y} \|z_1 + z_2\|_X \\ &\leq \inf_{z_1 \in x_1+Y} \|z_1\|_X + \inf_{z_2 \in x_2+Y} \|z_2\|_X \\ &= \|x_1+Y\|_{X/Y} + \|x_2+Y\|_{X/Y}. \end{aligned}$$

On the other hand, suppose that Y is not closed. Let x be a point that is not in Y but is in the closure of Y with respect to the distance $\|x_1 - x_2\|_X$. If $\{y_j\}_{j=1}^\infty$ is a sequence of points of Y that converges to x , then

$$0 \leq \|x+Y\|_{X/Y} = \inf_{z \in x+Y} \|z\|_X \leq \inf_{j \in \mathbf{N}} \|x - y_j\|_X = 0,$$

and consequently, $x \in Y$ which contradicts $x \notin Y$. Therefore $\|\cdot\|_{X/Y}$ could not be a norm on X/Y . We are done. \square

Example 8.1.14.

(i) If $X = \mathbf{R}^2$ and $Y = \{(0, y) : y \in \mathbf{R}\}$, then X/Y consists of lines in X of the form $(s, t) + Y$. Furthermore, if X is equipped with the norm $\|\cdot\|_{2,2}$, then $\|(s, t) + Y\|_{X/Y} = \inf_{(u, v) \in (s, t) + Y} \sqrt{u^2 + v^2} = |s|$.

(ii) The quotient space may be a little odd. For instance, let $\ell_{\infty,c}(\mathbf{N}, \mathbf{F})$ denote the space of all sequences $(x_j)_{j=1}^{\infty} \in \ell_{\infty}(\mathbf{N}, \mathbf{F})$ with the property that $\lim_{j \rightarrow \infty} x_j$ exists. This is a closed linear subspace of $\ell_{\infty}(\mathbf{N}, \mathbf{F})$. It is a good practice to give a representation of the quotient $\ell_{\infty}(\mathbf{N}, \mathbf{F}) / \ell_{\infty,c}(\mathbf{N}, \mathbf{F})$.

(iii) Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing, $E \subseteq \mathbf{R}$ an m_g -measurable set, $p \in [1, \infty]$. Then each set $LRS_g^p(E, \mathbf{F})$ defined in Example 6.1.3 (iii), is a linear space. If

$$N_g(E) = \{f : E \rightarrow \mathbf{F} : f = 0 \text{ } m_g - \text{a.e. on } E\},$$

then $LRS_g^p(E, \mathbf{F}) = LRS_g^p(E, \mathbf{F}) / N_g(E)$ forms a linear quotient space whose elements can be written as $f + N_g(E)$ where $f \in LRS_g^p(E, \mathbf{F})$. In particular, if $g : \mathbf{R} \rightarrow \mathbf{R}$ is the identity, then $LRS_{id}^p(E, \mathbf{F})$ is employed to represent $LRS_g^p(E, \mathbf{F})$. More importantly, when $1 \leq p \leq \infty$ the space $LRS_g^p(E, \mathbf{F})$ is a normed linear space under the following norm:

$$\|f + N_g(E)\|_{LRS_g^p(E, \mathbf{F})} = \|f\|_{p, m_g, E}.$$

The three properties required by a norm can be easily checked – in particular, the triangle inequality follows from Minkowski's inequality presented in Remark 4.5.7. Even though $LRS_g^p(E, \mathbf{F})$ exists as a space of equivalence classes of m_g -a.e.-defined $LRS_g^p(E, \mathbf{F})$ -functions, for the sake of simplicity of notation, we will follow the customary language to speak of it as a space of functions – this minor abuse of language is uniformly accepted and almost never results in any confusion.

8.2 Finite Dimensional Spaces

From the last section we have seen that any normed linear space generates a metric space and consequently brings the notions such as continuity and convergence into play. Therefore, it is unnecessary to deal with this issue again. But, to put it more abstractly, we want to emphasize that every normed linear space is a topological vector space and thus carries a topological structure which is induced by the norm and a linear mapping.

Definition 8.2.1.

(i) Two linear spaces X_1 and X_2 over \mathbf{F} are called algebraically isomorphic provided that there is a bijection $T : X_1 \rightarrow X_2$ that is linear:

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y) \text{ for all } \alpha, \beta \in \mathbf{F} \text{ and } x, y \in X_1.$$

(ii) A pair $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ of normed linear spaces over \mathbf{F} are called topologically isomorphic provided there is a linear bijection $T : X \rightarrow Y$ with the property that there exist two positive constants c_1 and c_2 independent of $x \in X$ with

$$c_1\|x\|_X \leq \|T(x)\|_Y \leq c_2\|x\|_X.$$

If $c_1 = c_2 = 1$ then T is called an isometry and the normed spaces X and Y are called isometric.

(iii) $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ defined on a given linear normed space X are equivalent provided that there are two constants $c_1, c_2 > 0$ such that

$$c_1\|\cdot\|_{(1)} \leq \|\cdot\|_{(2)} \leq c_2\|\cdot\|_{(1)}.$$

Example 8.2.2.

(i) Let Y be the set of all real polynomials of the form $f(t) = y_1 + y_2t + 2^{-1}y_3t^2$ with the norm $\|f\|_Y = \max\{|y_1|, |y_2|, |y_3|\}$, and X the set of all vectors $x = (x_1, x_2, x_3) \in \mathbf{R}^3$ with the norm $\|x\|_X = \sqrt{x_1^2 + x_2^2 + x_3^2}$. If $T : X \rightarrow Y$ is given by $T(x) = x_1 + x_2t + 2^{-1}x_3t^2$, then T is linearly bijective but also enjoys

$$\|T(x)\|_Y = \max\{|x_1|, |x_2|, |x_3|\} \leq \sqrt{|x_1|^2 + |x_2|^2 + |x_3|^2} = \|x\|_X$$

and

$$\sqrt{3}^{-1}\|x\|_X \leq \|T(x)\|_Y \leq \|x\|_X.$$

So, the two spaces are topologically isomorphic.

(ii) Given $p \in [1, \infty]$ and $n \in \mathbf{N}$, then the previously-defined norm $\|\cdot\|_{n,p}$ and the following norm $\|\cdot\|_{n,*}$ defined via

$$\|x\|_{n,*} = \max_{1 \leq k \leq n} |x_k| \text{ for } x = (x_1, \dots, x_n) \in \mathbf{R}^n,$$

are equivalent.

In fact, we have the following more general result.

Theorem 8.2.3. Let $n \in \mathbf{N}$.

(i) Any two norms on a given n -dimensional linear space X over \mathbf{F} are equivalent.

(ii) If X and Y are n -dimensional normed linear spaces over \mathbf{F} then X and Y are topologically isomorphic.

Proof. (i) Let $\|\cdot\|_{(k)}$, $k = 1, 2$, be two norms on X . Choose a basis $\{e_1, \dots, e_n\}$ for X and define a third norm $\|\cdot\|_{(3)}$ as follows. For each $x \in X$ there is a unique set of scalars $\alpha_1, \dots, \alpha_n$ in \mathbf{F} such that $x = \sum_{j=1}^n \alpha_j e_j$. Let $\|x\|_{(3)} = \sum_{j=1}^n |\alpha_j|$. Suppose each of the norms $\|\cdot\|_{(k)}$, $k = 1, 2$ is equivalent to $\|\cdot\|_{(3)}$. Then there are positive constants c_1, C_1 and c_2, C_2 such that

$$c_k \|\cdot\|_{(k)} \leq \|\cdot\|_{(3)} \leq C_k \|\cdot\|_{(k)} \quad \text{for } k = 1, 2.$$

It follows that

$$c_1 C_2^{-1} \|\cdot\|_{(1)} \leq \|\cdot\|_{(2)} \leq C_1 c_2^{-1} \|\cdot\|_{(1)},$$

implying the desired equivalence.

Now let $\|\cdot\|$ denote either $\|\cdot\|_{(1)}$ or $\|\cdot\|_{(2)}$. We show that $\|\cdot\|$ is equivalent to $\|\cdot\|_{(3)}$. If $x = \sum_{j=1}^n \alpha_j e_j$ then

$$\|x\| \leq \sum_{j=1}^n |\alpha_j| \|e_j\| \leq \left(\max_{1 \leq j \leq n} \|e_j\| \right) \|x\|_{(3)}.$$

This implies a one-sided estimate. We shall prove that there is a constant $c > 0$ such that $\|\cdot\|_{(3)} \leq c \|\cdot\|$, i.e.,

$$\sum_{j=1}^n |\alpha_j| \leq c \left\| \sum_{j=1}^n \alpha_j e_j \right\|.$$

We may assume that $\|x\|_{(3)} = 1$ by dividing the last inequality by $\sum_{j=1}^n |\alpha_j|$. Now if the last inequality is not true then there would be a sequence $\{y_k\}_{k=1}^{\infty}$ where

$$y_k = \sum_{j=1}^n \alpha_{k,j} e_j \quad \text{with} \quad \|y_k\|_{(3)} = \sum_{j=1}^n |\alpha_{k,j}| = 1 \quad \text{but} \quad \lim_{k \rightarrow \infty} \|y_k\| = 0.$$

Note that $|\alpha_{k,j}| \leq 1$ for $j = 1, \dots, n$ and every k . So there is a subsequence of $\{y_k\}_{k=1}^{\infty}$, denoted itself also, such that $\lim_{k \rightarrow \infty} \alpha_{k,j}$ exists and equals, say, α_j for $j = 1, 2, \dots, n$ as well as

$$\lim_{k \rightarrow \infty} \left\| y_k - \sum_{j=1}^n \alpha_j e_j \right\| \leq \left(\max_{1 \leq j \leq n} \|e_j\| \right) \lim_{k \rightarrow \infty} \sum_{j=1}^n |\alpha_{k,j} - \alpha_j| = 0.$$

Consequently,

$$\sum_{j=1}^n |\alpha_j| = 1 \quad \text{while} \quad \sum_{j=1}^n \alpha_j e_j = 0.$$

This is impossible, since $\{e_j\}_{j=1}^n$ are linearly independent. This completes the argument.

(ii) It suffices to show that X is topologically isomorphic to \mathbf{F}^n with the following Euclidean norm:

$$\|\alpha\|_{n,2} = \left(\sum_{k=1}^n |\alpha_k|^2 \right)^{2^{-1}} \quad \text{for } \alpha = (\alpha_1, \dots, \alpha_n) \in \mathbf{F}^n.$$

By the argument for (i), there is an isomorphism T from X onto \mathbf{F}^n . We can use this map to define a new norm on X as follows: For each $x \in X$ let $\|x\|_X = \|T(x)\|_{n,2}$. When X is given this new norm, T becomes a topological isomorphism. Nevertheless, From (i) it turns out that $\|\cdot\|_X$ is equivalent to any norm $\|\cdot\|_X$ on X . Hence T is a topological isomorphism from $(X, \|\cdot\|_X)$ onto \mathbf{F}^n . \square

Interestingly, Theorem 8.2.3 has some important consequences.

Corollary 8.2.4. *Let $n \in \mathbf{N}$ and X be a normed linear space over \mathbf{F} .*

- (i) *If X is n -dimensional, then it is complete.*
- (ii) *Any n -dimensional linear subspace of X is closed.*
- (iii) *If X is n -dimensional, then every bounded closed subset S of X is compact.*

Proof. (i) Let $\{e_k\}_{k=1}^n$ be a basis for X . Then each $x \in X$ can be written as $x = \sum_{j=1}^n \alpha_j e_j$, $\alpha_j \in \mathbf{F}$. Set $\|x\|_{n,2} = (\sum_{j=1}^n |\alpha_j|^2)^{\frac{1}{2}}$. This is a norm on X , and by Theorem 8.2.3 it is equivalent to the given norm on X . Accordingly, if $\{x^{(k)} = \sum_{j=1}^n \alpha_j^{(k)} e_j\}_{k=1}^\infty$ is a Cauchy sequence in X , then $\{(\alpha_1^{(k)}, \dots, \alpha_n^{(k)})\}_{k=1}^\infty$ is a Cauchy sequence in \mathbf{F}^n which is complete, and hence there exists a point $(\alpha_1, \dots, \alpha_n) \in \mathbf{F}^n$ such that

$$(\alpha_1^{(k)}, \dots, \alpha_n^{(k)}) \rightarrow (\alpha_1, \dots, \alpha_n) \quad \text{as } k \rightarrow \infty.$$

Therefore $\{x^{(k)}\}_{k=1}^\infty$ is convergent to $x = \sum_{j=1}^n \alpha_j e_j$ in X , whence X is complete.

(ii) Suppose that S is an n -dimensional linear subspace of X . Then S is complete due to (i) and hence any sequence $\{x^{(k)}\}_{k=1}^\infty$ in S which is convergent to $x \in X$ must be a Cauchy sequence in S – this yields $x \in S$ and so S is closed.

(iii) By Theorem 8.2.3 (ii) it follows that X is topologically isomorphic to \mathbf{F}^n and hence every bounded closed subset S of X is topologically isomorphic to a compact (i.e., bounded and closed) subset of \mathbf{F}^n . Accordingly, S is compact. \square

Even more interestingly, we can say something more about the converse of Corollary 8.2.4 (iii).

Theorem 8.2.5. *Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} .*

(i) *If S is a proper closed subspace of X , then for each number $\theta \in (0, 1)$ there exists an element $x_\theta \in X$ such that*

$$\|x_\theta\|_X = 1 \text{ and } d(x_\theta, S) = \inf_{x \in S} \|x_\theta - x\|_X \geq \theta.$$

(ii) *If the unit sphere $\{x \in X : \|x\|_X = 1\}$ of X is compact, then X is finite dimensional.*

Proof. (i) Since S is closed and not equal to X , there exists an $x_1 \in X \setminus S$ such that $\delta = \inf_{x \in S} \|x_1 - x\|_X > 0$. Consequently, for any $\varepsilon > 0$ there is an $x_2 \in S$ such that $\|x_2 - x_1\|_X < \delta + \varepsilon$. Choosing

$$\varepsilon = \delta(\theta^{-1} - 1) \text{ and } x_\theta = x_2 - x_1 \|x_2 - x_1\|_X^{-1},$$

we obtain $\|x_\theta\|_X = 1$ and

$$\|x - x_\theta\|_X = \|(\|x_2 - x_1\|_X x + x_1) - x_2\|_X \|x_2 - x_1\|_X^{-1} \geq \delta \|x_2 - x_1\|_X^{-1} \geq \theta$$

for any $x \in S$. Here we have used the fact that S is a linear subspace of X .

(ii) Suppose that x_1 is any element of X with $\|x_1\|_X = 1$ and S_1 is the subspace spanned by x_1 . If $S_1 = X$ then X is finite dimensional and we are done. Otherwise, by Theorem 8.2.5 (i), Corollary 8.2.4 (ii) and the fact that S_1 is finite dimensional, there exists an $x_2 \in X$ such that $\|x_2\|_X = 1$ and $d(x_2, S_1) \geq 2^{-1}$. Suppose S_2 is the linear subspace of X spanned by x_1 and x_2 . Then S_2 is a closed linear subspace of X and consequently, if $S_2 \neq X$ then by Theorem 8.2.5 (i) there exists an $x_3 \in X$ such that $\|x_3\|_X = 1$ and $d(x_3, S_2) \geq 2^{-1}$. Continuing in this manner, we find inductively that if the closed linear subspace S_n spanned by x_1, \dots, x_n is not equal to X then by Theorem 8.2.5 (i) there exists an $x_{n+1} \in X$ such that $\|x_{n+1}\|_X = 1$ and $d(x_{n+1}, S_n) \geq 2^{-1}$. However, this process must stop at some step – for otherwise there would be an infinite sequence $\{x_j\}_{j=1}^\infty$ with $\|x_j\|_X = 1$ but $\|x_j - x_k\|_X \geq 2^{-1}$ when $j \neq k$. Of course, this sequence has no

convergent subsequence, thereby deriving that the unit sphere of X is not compact (thanks to Theorem 6.3.7 (iv)), a contradiction. Accordingly, X equals some S_k , and the argument is complete. \square

The foregoing discussion reveals that the unit sphere of a given normed linear space is compact if and only if the space is finite dimensional, but also leads naturally to the following concept.

Definition 8.2.6. *If $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are normed linear spaces over \mathbf{F} , then any one of:*

- (i) $\|(x, y)\| = (\|x\|_X^p + \|y\|_Y^p)^{\frac{1}{p}}$ under $p \in [1, \infty)$;
- (ii) $\|(x, y)\| = \max\{\|x\|_X, \|y\|_Y\}$.

may be defined as a norm on the Cartesian product $X \times Y$.

This certainly does not exhaust all the possible combinations of the norms $\|x\|_X$ and $\|y\|_Y$, and yet these are the most commonly used ones. An extension to the Cartesian product of n normed linear spaces is defined in the following manner.

Remark 8.2.7. *For $n - 1 \in \mathbf{N}$ let $\{X_k\}_{k=1}^n$ be a collection of linear spaces over \mathbf{F} with norms $\{\|\cdot\|_{X_k}\}_{k=1}^n$. If their Cartesian product space is defined as*

$$\prod_{k=1}^n X_k = X_1 \times \cdots \times X_n = \{(x_1, \dots, x_n) : x_1 \in X_1, \dots, x_n \in X_n\},$$

with vector addition

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$$

and scalar multiplication

$$\alpha(x_1, \dots, x_n) = (\alpha x_1, \dots, \alpha x_n)$$

for

$$\alpha \in \mathbf{F} \text{ and } (x_1, \dots, x_n), (y_1, \dots, y_n) \in \prod_{k=1}^n X_k.$$

Then it is not hard to show that the function $\prod_{k=1}^n X_k \rightarrow \mathbf{R}$ defined by

$$\|(x_1, \dots, x_n)\|_{\prod_{k=1}^n X_k} = \sum_{k=1}^n \|x_k\|_{X_k}$$

is a norm on X .

Example 8.2.8. If $n = m + k$ with $k, m, n \in \mathbf{N}$, then \mathbf{R}^n may be viewed as the Cartesian product of \mathbf{R}^m and \mathbf{R}^k .

8.3 Bounded Linear Operators

In this section, we discuss bounded operators, or equivalently, continuous operators, and inverse operators. To begin with, we give the definition for a linear operator to be bounded.

Definition 8.3.1. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces over \mathbf{F} . A linear operator $T : X \rightarrow Y$ is said to be bounded provided there exists a constant $C \geq 0$ such that $\|T(x)\|_Y \leq C\|x\|_X$ for all $x \in X$. Define

$$\|T\| = \|T\|_{X \rightarrow Y} = \sup_{x \in X, x \neq 0} \|T(x)\|_Y \|x\|_X^{-1}.$$

According to this definition, we have $\|T\| = \sup_{\|x\|_X=1} \|T(x)\|_Y$ which follows from

$$\|x\|_X \|x\|_X^{-1} \|T(x)\|_Y = \|T(x)\|_Y \text{ for all } x \neq 0.$$

Example 8.3.2.

(i) Equip ℓ_2 with the 2-norm, and let $T : \ell_2 \rightarrow \ell_2$ be given by $T(x) = (0, x_1, x_2, \dots)$ when $x = (x_1, x_2, \dots)$. Then it is easy to check that T is a bounded linear operator with $\|T\| = 1$.

(ii) Choose for both X and Y the real space $C[0, 1]$ with the sup-norm. Define $T : X \rightarrow Y$ by $T(f)(x) = e^x f(x)$, $x \in [0, 1]$. Then T is bounded with $\|T\| = e$.

The following result tells us that as for a linear operator, the boundedness amounts to the continuity.

Theorem 8.3.3. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces over \mathbf{F} and $T : X \rightarrow Y$ be a linear operator. Then the following statements are equivalent:

- (i) T is continuous in X ;
- (ii) T is continuous at 0;
- (iii) T is bounded;
- (iv) T maps bounded subsets of X to bounded subsets of Y .

Proof. In the sequel, we will write $A \Leftrightarrow B$ for the statement: If A then B and vice versa, but also use the criterion for a linear map $T : X \rightarrow Y$ to be continuous at a point $x \in X$:

$$\|x_n - x\|_X \rightarrow 0 \Rightarrow \|T(x_n) - T(x)\|_Y = 0.$$

(i) \Leftrightarrow (ii) The implication (i) \Rightarrow (ii) is trivial. Regarding (ii) \Rightarrow (i), we suppose that T is continuous at $0 \in X$. If $x_n \rightarrow x$ then $x_n - x \rightarrow 0$. Hence

$$T(x_n - x) \rightarrow T(0) = 0 \text{ so that } T(x_n) \rightarrow T(x) \text{ in } Y.$$

(iii) \Leftrightarrow (i) If T is bounded and $x_n \rightarrow x$ in X , then $T(x_n) \rightarrow T(x)$ as well. It follows that T is continuous at x . Conversely, assume that T is continuous in X . If T is not bounded, then for any $n \in \mathbf{N}$ there exists a point $x_n \in X$ with $\|T(x_n)\|_Y \geq n\|x_n\|_X$. Let $y_n = x_n(n\|x_n\|_X)^{-1}$, so that $\|y_n\|_X = n^{-1} \rightarrow 0$. However, $\|T(y_n)\|_Y > 1$ and $T(0) = 0$, contradicting the assumption that T is continuous at 0 .

(iv) \Leftrightarrow (ii) Suppose that (ii) is true. So T is bounded with $\|T\| \leq C_1$ for some constant $C_1 > 0$ due to (iii) \Leftrightarrow (ii). If S is a bounded subset of X , then there is a constant $C_2 > 0$ such that $\|x\|_X \leq C_2$ for all $x \in S$, and hence

$$\|T(x)\|_Y \leq C_1\|x\|_X \leq C_1C_2 \text{ for all } x \in X.$$

That is to say, $T(S)$ is a bounded subset of Y . Conversely, assume that (iv) is true. Given an open ball

$$B_\varepsilon^Y(0) = \{y \in Y : \|y\|_Y < \varepsilon\} \text{ in } Y$$

let

$$B_1^X(0) = \{x \in X : \|x\|_X < 1\}$$

denote the open unit ball in X . By (iv) it follows that $T(B_1^X)$ is bounded in Y . Thus, there is a $\lambda > 0$ such that

$$T(B_1^X(0)) \subseteq \lambda B_\varepsilon^Y(0) = \{y \in Y : \|y\|_Y < \lambda\varepsilon\}.$$

This implies $T(B_{\lambda^{-1}}^X(0)) \subseteq B_\varepsilon^Y$ since T is linear, and so T is continuous at 0 . \square

Example 8.3.4. For $n \in \mathbf{N}$ equip $X = \mathbf{R}^n$ with the $(n, 2)$ -norm and let $\{e_j\}_{j=1}^n$ (for which the j -th coordinate of e_j is 1 and others are 0) be the standard basis for \mathbf{R}^n . Then any $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ has the form

$$x = \sum_{j=1}^n x_j e_j \text{ with } \|x\|_2^2 = \sum_{j=1}^n |x_j|^2.$$

If $T : \mathbf{R}^n \rightarrow Y$ is a linear transformation, where $(Y, \|\cdot\|_Y)$ is a normed linear space over \mathbf{F} , then the Cauchy-Schwarz inequality implies

$$\|T(x)\|_Y \leq \|x\|_2 \left(\sum_{j=1}^n \|T(e_j)\|_Y^2 \right)^{2^{-1}}.$$

So, T is bounded with

$$\|T\| \leq \left(\sum_{j=1}^n \|T(e_j)\|_Y^2 \right)^{2^{-1}}.$$

Indeed, Example 8.3.4 is a special case of the following result.

Theorem 8.3.5. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed linear spaces over \mathbf{F} . If X is finite dimensional, then any linear transformation $T : X \rightarrow Y$ is bounded.

Proof. Note that any two norms on a finite dimensional linear space X over \mathbf{F} are equivalent. So, we construct a new norm $\|\cdot\|$ via $\|\cdot\|_X$ and $\|\cdot\|_Y$:

$$\|x\| = \|x\|_X + \|T(x)\|_Y \text{ for all } x \in X.$$

Of course, $\|\cdot\|$ is a norm on X and so it is equivalent to $\|\cdot\|_X$. This yields a constant $C > 0$ with $\|\cdot\| \leq C\|\cdot\|_X$. It follows that $\|T(x)\|_Y \leq \|x\| \leq C\|x\|_X$ for all $x \in X$. In other words, $T : X \rightarrow Y$ is bounded. \square

Next, we consider some basic algebraic properties of the space of all bounded linear operators.

Definition 8.3.6. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed linear spaces over \mathbf{F} . Denote by $\mathcal{B}(X, Y)$ the set of all bounded linear operators from X to Y . In particular, $\mathcal{B}(X) = \mathcal{B}(X, X)$.

Theorem 8.3.7. Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ and $(Z, \|\cdot\|_Z)$ be normed linear spaces over \mathbf{F} . Then

(i) $\mathcal{B}(X, Y)$ is a linear space over \mathbf{F} with respect to the operations for $T, S \in \mathcal{B}(X, Y)$:

$$(T + S)(x) = T(x) + S(x) \text{ and } (\alpha T)(x) = \alpha T(x) \text{ for } x \in X, \alpha \in \mathbf{F}.$$

(ii) The function $\|\cdot\| : \mathcal{B}(X, Y) \rightarrow \mathbf{R}$, defined by

$$\|T\| = \sup_{x \in X, \|x\|_X \neq 0} \|T(x)\|_Y \|x\|_X^{-1} \text{ for } T \in \mathcal{B}(X, Y),$$

is a norm on $\mathcal{B}(X, Y)$.

(iii) If $T \in \mathcal{B}(X, Y)$ and $S \in \mathcal{B}(Y, Z)$, then the composition $ST = S \circ T$ belongs to $\mathcal{B}(X, Z)$ with $\|ST\| \leq \|S\| \|T\|$.

(iv) If $T, S \in \mathcal{B}(X)$ are invertible; that is, there are two elements $T^{-1}, S^{-1} \in \mathcal{B}(X)$ such that

$$TT^{-1} = T^{-1}T = I \text{ and } SS^{-1} = S^{-1}S = I,$$

where I stands for the identity element of $\mathcal{B}(X)$, then $ST \in \mathcal{B}(X)$ is invertible with $(ST)^{-1} = T^{-1}S^{-1}$.

Proof. (i) This follows from checking those conditions for a linear space with $\mathcal{B}(X, Y)$.

(ii) We have to verify three conditions required for a norm. First, it is clear that $\|T\| \geq 0$. If $\|T\| = 0$ then $\|T(x)\|_Y = 0$ for all $x \in X$, and hence $T(x) = 0$ for all $x \in X$. This gives $T = 0$. Conversely, $T = 0$ implies $\|T\| = 0$.

Next,

$$\|\alpha T\| = \sup_{x \in X, \|x\|_X \neq 0} \|\alpha T(x)\|_Y \|x\|_X^{-1} = |\alpha| \|T\|.$$

Finally,

$$\begin{aligned} \|T + S\| &= \sup_{x \in X, \|x\|_X \neq 0} \|T(x) + S(x)\|_Y \|x\|_X^{-1} \\ &\leq \sup_{x \in X, \|x\|_X \neq 0} \|T(x)\|_Y \|x\|_X^{-1} + \sup_{x \in X, \|x\|_X^{-1} \neq 0} \|S(x)\|_Y \|x\|_X^{-1} \\ &= \|T\| + \|S\|. \end{aligned}$$

(iii) Clearly, ST is a linear transformation from X to Z . Since T and S are bounded, we conclude

$$\|(ST)x\|_Z = \|S(T(x))\|_Z \leq \|S\| \|T(x)\|_Y \leq \|S\| \|T\| \|x\|_X \quad \text{for all } x \in X.$$

This yields that ST is bounded and so in $\mathcal{B}(X, Z)$ with $\|ST\| \leq \|S\| \|T\|$, as desired.

(iv) This follows from the following formula:

$$(ST)T^{-1}S^{-1} = I = T^{-1}S^{-1}(ST).$$

□

Example 8.3.8. Define $T : \ell_2 \rightarrow \ell_2$ by $T((x_1, x_2, \dots)) = (0, x_1, x_2, \dots)$. This map is bounded but not invertible, since it is clearly not onto. If $S : \ell_2 \rightarrow \ell_2$ is given by $S((x_1, x_2, \dots)) = (x_2, x_3, \dots)$, then it is bounded but not invertible, because it is clearly not one to one. Note that $ST = I \neq TS$.

8.4 Linear Functionals via Hahn-Banach Extension

Looking at the Lebesgue-Radon-Stieltjes integrals, we find that if $g : \mathbf{R} \rightarrow \mathbf{R}$ is increasing then $L_g(f) = \int_{\mathbf{R}} f dm_g$ defines a linear function from the real linear space $LRS_g(\mathbf{R})$ into \mathbf{R} . This observation suggests that we should work with the so-called linear functionals.

Definition 8.4.1. Let X be a linear space over \mathbf{F} . A linear map from X to \mathbf{F} is called an \mathbf{F} -valued linear functional on X . If X is a normed linear space, then $X^* = \mathcal{B}(X, \mathbf{F})$ and $X^{**} = (X^*)^*$ are respectively called the dual space and second dual space of X over \mathbf{F} ; moreover X is said to be reflexive whenever X^{**} and X are isometric.

The question is whether or not the class of \mathbf{F} -valued linear functionals on X consists only of the zero functional. To settle this question, let us consider the following example of Mahlon Marsh Day type.

Example 8.4.2. Let $p \in (0, 1)$ and $I = [0, 1]$. If L is a continuous linear function on $LRS_{id}^p(I, \mathbf{R})$ then L must be the zero functional.

To see this conclusion, suppose that, to the contrary, $L \neq 0$. Then, without loss of generality we may assume that there exists an $f \in LRS_{id}^p(I, \mathbf{R})$ such that

$$L(f) = 1 \quad \text{and} \quad \|f\|_{p, m_{id}, I} \neq 0.$$

Hence, $f1_{[0,x]}$ is continuous in the metric of $LRS_{id}^p(I, \mathbf{R})$, and consequently, $\phi(x) = L(f1_{[0,x]})$ is a continuous function on I since L is continuous. Note that

$$\begin{cases} \phi(0) = 0; \\ \phi(1) = 1; \\ f = f1_{[0,x]} + f1_{[x,1]}. \end{cases}$$

So there is an $x_0 \in (0, 1)$ such that

$$\phi(x_0) = 2^{-1} \text{ and } L(f1_{[0,x_0]}) = 2^{-1} = L(f1_{[x_0,1]}).$$

Since

$$\|f1_{[0,x_0]}\|_{p,m_{id},I}^p + \|f1_{[x_0,1]}\|_{p,m_{id},I}^p = \|f\|_{p,m_{id},I}^p,$$

one of two left terms of the last equation is not greater than $2^{-1}\|f\|_{p,m_{id},I}^p$, for example,

$$\|f1_{[0,x_0]}\|_{p,m_{id},I}^p \leq 2^{-1}\|f\|_{p,m_{id},I}^p$$

and then

$$L(2f1_{[0,x_0]}) = 1 \text{ and } \|2f1_{[0,x_0]}\|_{p,m_{id},I}^p \leq 2^{p-1}\|f\|_{p,m_{id},I}^p.$$

Continuing the above argument, we obtain a sequence $\{f_j\}_{j=1}^\infty$ in $LRS_{id}^p(I, \mathbf{R})$ such that for any $j \in \mathbf{N}$,

$$L(f_j) = 1 \text{ and } \|f_j\|_{p,m_{id},I}^p \leq 2^{j(p-1)}\|f\|_{p,m_{id},I}^p.$$

But, the last inequality cannot hold due to $0 < p < 1$, $2^{j(p-1)} \rightarrow 0$, and L being continuous.

Actually, the above question will be answered in great generality using the Hahn-Banach extension theorem which is stated below and named for Hans Hahn and Stefan Banach.

Theorem 8.4.3. Let X be a linear space over \mathbf{F} , and $p : X \rightarrow [0, \infty)$ a function with

$$p(x+y) \leq p(x) + p(y) \text{ and } p(\lambda x) = |\lambda|p(x) \text{ for all } \lambda \in \mathbf{F} \text{ and } x, y \in X.$$

If Y is a linear subspace of X and f is an \mathbf{F} -valued linear functional on Y with $|f(x)| \leq p(x)$ for all $x \in Y$, then there is an \mathbf{F} -valued linear functional F on X such that $F(x) = f(x)$ for all $x \in Y$ and $|F(x)| \leq p(x)$ for all $x \in X$.

Proof. First of all, we demonstrate the theorem for $\mathbf{F} = \mathbf{R}$. By the hypothesis, we denote by \mathcal{K} the set of all pairs (Y_α, g_α) in which Y_α is a linear subspace of X containing Y , and g_α is a real linear functional on Y_α with

$$g_\alpha(x) = f(x) \text{ for all } x \in Y \text{ and } g_\alpha(x) \leq p(x) \text{ for all } x \in Y_\alpha.$$

Make \mathcal{K} into a partially ordered set by defining the relation:

$$(Y_\alpha, g_\alpha) \preceq (Y_\beta, g_\beta) \text{ if } Y_\alpha \subseteq Y_\beta \text{ and } g_\alpha = g_\beta \text{ on } Y_\alpha.$$

Clearly, every totally ordered subset $\{(Y_\lambda, g_\lambda)\}$ of \mathcal{K} , for which at least one of

$$(Y_\alpha, g_\alpha) \preceq (Y_\beta, g_\beta) \text{ and } (Y_\beta, g_\beta) \preceq (Y_\alpha, g_\alpha)$$

holds, has an upper bound $\cup_\lambda Y_\lambda$ on which the functional is given by g_λ on each Y_λ . By Zorn's lemma we find that there is a maximal element (Y_0, g_0) in \mathcal{K} . Subsequently, we will prove $Y_0 = X$ and $g_0 = F$.

If $Y_0 \neq X$, then there is a $y_1 \in X \setminus Y_0$. Let Y_1 be the linear space spanned by Y_0 and y_1 ; that is,

$$Y_1 = \{y + \lambda y_1 : y \in Y_0, \lambda \in \mathbf{R}\}.$$

Note that Y_0 is a subset of Y_1 and that

$$\begin{aligned} x, y \in Y_0 &\Rightarrow g_0(y) - g_0(x) = g_0(y - x) \leq p(y - x) \leq p(y + y_1) + p(-y_1 - x) \\ &\Rightarrow -p(-y_1 - x) - g_0(x) \leq p(y + y_1) - g_0(y). \end{aligned}$$

So, it follows that

$$a = \sup_{x \in Y_0} \{-p(-y_1 - x) - g_0(x)\} \leq \inf_{y \in Y_0} \{p(y + y_1) - g_0(y)\} = b.$$

Now for any number $c \in [a, b]$ define

$$g_1(y + \lambda y_1) = g_0(y) + \lambda c \text{ for all } (y, \lambda) \in Y_0 \times \mathbf{R}.$$

Then g_1 is evidently linear. Furthermore, we have the following three cases:

Case 1: If $\lambda = 0$, then $g_1(y) = g_0(y) \leq p(y)$.

Case 2: If $\lambda > 0$ and $y \in Y_0$, then

$$\begin{aligned} g_1(y + \lambda y_1) &= \lambda(g_0(\lambda^{-1}y) + c) \\ &\leq \lambda(g_0(\lambda^{-1}y) + p(\lambda^{-1}y + y_1) - g_0(\lambda^{-1}y)) \\ &= \lambda p(\lambda^{-1}y + y_1) \\ &= p(y + \lambda y_1), \end{aligned}$$

Case 3: If $\lambda < 0$, then

$$\begin{aligned}
 g_1(y + \lambda y_1) &= |\lambda| (g_0(|\lambda|^{-1}y) - c) \\
 &\leq |\lambda| (g_0(|\lambda|^{-1}y) - g_0(|\lambda|^{-1}y) + p(-y_1 + |\lambda|^{-1}y)) \\
 &= |\lambda| p(-y_1 + |\lambda|^{-1}y) \\
 &= p(y + \lambda y_1).
 \end{aligned}$$

Accordingly, we get

$$g_1(y + \lambda y_1) = g_0(y) + \lambda c \leq p(y + \lambda y_1) \quad \text{for all } (\lambda, y) \in \mathbf{R} \times Y_0.$$

This is to say,

$$(Y_1, g_1) \in \mathcal{K} \quad \text{and} \quad (Y_0, g_0) \preceq (Y_1, g_1) \quad \text{with } Y_0 \neq Y_1.$$

This contradicts the maximality of (Y_0, g_0) . The preceding argument indicates that if $x \in X$ and $\lambda = \text{sgn}F(x)$, the real number obeying $\lambda F(x) = |F(x)|$, then

$$|F(x)| = \lambda F(x) = F(\lambda x) \leq p(\lambda x) \leq p(x).$$

Next, we prove the theorem for the case $\mathbf{F} = \mathbf{C}$. Let f be a complex linear functional on Y such that $|f(x)| \leq p(x)$ for $x \in Y$. Then $u = \Re f$ is clearly real linear on Y and $|u(x)| = |\Re f(x)| \leq |f(x)| \leq p(x)$. According to the above \mathbf{R} -case there is a real-valued linear functional U on X such that $U(x) = u(x)$ for $x \in Y$ and $|U(x)| \leq p(x)$ for $x \in X$. Now if $F(x) = U(x) - iU(ix)$, then

$$\begin{cases} \Im F(x) = -\Re(iF(x)) = -U(ix); \\ F(ix) = U(ix) - iU(-x) = U(ix) + iU(x) = iF(x), \end{cases}$$

and consequently, F is a complex linear extension of $f(x) = u(x) - iu(ix)$ to X . Furthermore, if

$$\alpha = \text{sgn}F(x) = \begin{cases} \exp(-i \arg F(x)), & F(x) \neq 0 \\ 0, & F(x) = 0 \end{cases},$$

then

$$|F(x)| = \alpha F(x) = F(\alpha x) = U(\alpha x) \leq p(\alpha x) \leq p(x).$$

□

Here is a series of important and interesting consequences of the Hahn-Banach extension theorem applied to normed linear spaces.

Corollary 8.4.4. *Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} and Y a linear subspace of X .*

(i) *To any $f \in Y^*$ there corresponds an $F \in X^*$ such that*

$$\|F\| = \|f\| \text{ and } F(y) = f(y) \text{ for all } y \in Y.$$

(ii) *If $x_0 \in X$ satisfies $\inf_{y \in Y} \|y - x_0\|_X = d > 0$, then there is an $F \in X^*$ such that*

$$F(x_0) = 1; \quad \|F\| = d^{-1}; \quad F(y) = 0 \text{ for all } y \in Y.$$

(iii) *If Y is not dense in X , then there is a nonzero $F \in X^*$ such that $F(y) = 0$ for all $y \in Y$.*

(iv) *If $x \neq 0$ in X then there is an $F \in X^*$ such that $\|F\| = 1$ and $F(x) = \|x\|_X$.*

(v) *If $y, z \in X$ and $y \neq z$, then there exists an $F \in X^*$ such that $F(y) \neq F(z)$.*

(vi) *For any $x \in X$,*

$$\|x\|_X = \sup_{0 \neq F \in X^*} |F(x)| \|F\|^{-1} = \sup_{F \in X^*, \|F\|=1} |F(x)|.$$

(vii) *If $F \in X^*$ is nonzero and $N = \{x \in X : F(x) = 0\}$, then there exists a one-dimensional subspace M of X such that*

$$X = N + M = \{x + y : x \in N \text{ and } y \in M\} \text{ and } N \cap M = \{0\}.$$

(viii) *If X^* is separable, so is X .*

(ix) *If Y is closed and $Y^\perp = \{f \in X^* : f(y) = 0 \text{ for all } y \in Y\}$, then $Y^* = X^*/Y^\perp$ and $(X/Y)^* = Y^\perp$ in the sense that there exist isometries between the corresponding normed spaces.*

(x) *If X is reflexive and Y is closed, then Y is reflexive.*

Proof. (i) Given $f \in Y^*$, let $p(x) = \|f\| \|x\|_X$ for all $x \in X$. Then

$$|f(x)| \leq \|f\| \|x\|_X = p(x) \text{ for all } x \in Y.$$

Hence from Theorem 8.4.3 it turns out that there exists an extension $F \in X^*$ with $F = f$ on Y and $|F| \leq p$ on X . Clearly, $\|F\| = \|f\|$ follows from $\|F\| \leq \|f\|$ and

$$\|f\| = \sup_{y \in Y, \|y\|_X=1} |f(y)| = \sup_{y \in Y, \|y\|_X=1} |F(y)| \leq \|F\|.$$

(ii) Let Y_1 be the linear space spanned by Y and x_0 . Since $d > 0$, we conclude that $x_0 \notin Y$ and every point $x \in Y_1$ may be written uniquely as $x = y + \lambda x_0$ where $y \in Y$ and $\lambda \in \mathbf{F}$. Define a linear functional

$$f \in Y_1^* \text{ by } f(y + \lambda x_0) = \lambda.$$

Then $f(y) = 0$ for $y \in Y$ and $f(x_0) = 1$. If $\lambda \neq 0$ and $x = y + \lambda x_0$, then

$$\|x\|_X = \|y + \lambda x_0\|_X = |\lambda| \|\lambda^{-1}y + x_0\|_X \geq |\lambda|d = |f(x)|d,$$

and hence $\|f\| \leq d^{-1}$. Pick a sequence $(y_n)_{n=1}^\infty$ in Y with $\|x_0 - y_n\|_X \rightarrow d$. Then

$$1 = f(x_0 - y_n) \leq \|f\| \|x_0 - y_n\|_X \rightarrow d\|f\|,$$

so $\|f\| \geq d^{-1}$. Therefore $\|f\| = d^{-1}$. Accordingly, a direct application of (i) produces an $F \in X^*$ such that $F(x) = f(x)$ when $x \in Y_1$, and $\|F\| = \|f\|$, as desired.

(iii) Since Y is not dense in X , we conclude that there is an $x_0 \in X$ such that $\inf_{y \in Y} \|y - x_0\|_X = d > 0$. An application of (ii) produces the conclusion in (iii).

(iv) Just apply (ii) with $Y = \{0\}$ and $x \in X$ to get an $f \in X^*$ such that $\|f\| = \|x\|_X^{-1}$ and $f(x) = 1$. We may then take $F = \|x\|_X f$.

(v) Apply (iv) to $x = y - z$.

(vi) If $x = 0$ then the assertion is trivial. So, suppose $x \neq 0$. Then

$$\sup_{F \in X^*, \|F\|=1} |F(x)| \leq \|x\|_X.$$

Meanwhile, by (iv) there is an $f \in X^*$ such that $f(x) = \|x\|_X$ and $\|f\| = 1$, thus

$$\sup_{F \in X^*, \|F\|=1} |F(x)| = \|x\|_X.$$

(vii) If $F \neq 0$ then there is a point $x_0 \neq 0$ such that $F(x_0) = 1$. Note that any element $x \in X$ can then be written as

$$x = x - \lambda x_0 + \lambda x_0 \text{ with } \lambda = F(x).$$

So, if $Y = \{\lambda x_0 : \lambda \in \mathbf{F}\}$ then the desired decomposition follows right away. It is clear that Y is the one-dimensional space spanned by x_0 . If $x \in N \cap Y$, then

$$x = \lambda x_0 \text{ and } 0 = F(x) = \lambda F(x_0) = \lambda$$

and hence $x = 0$.

(viii) Suppose that $\{f_j\}_{j=1}^\infty$ is a dense subset of the unit sphere $\{f \in X^* : \|f\| = 1\}$ and choose $x_j \in \{x \in X : \|x\|_X = 1\}$ to obey $|f_j(x_j)| \geq 2^{-1}$. Let $Y = \{x_j\}_{j=1}^\infty$. If $\bar{Y} \neq X$, then the just-verified (iii) gives a functional $f \in X^*$ such that $\|f\| = 1$ and $f(x) = 0$ for all $x \in Y$. Now if $\varepsilon \in (0, 2^{-1})$, then there exists a functional $f_{j_0} \in \{f_j\}_{j=1}^\infty$ such that $x_{j_0} \in Y$ and

$$2^{-1} \leq |f_{j_0}(x_{j_0})| = |f(x_{j_0}) - f_{j_0}(x_{j_0})| \leq \|f - f_{j_0}\| < \varepsilon,$$

a contradiction. Accordingly, $\bar{Y} = X$, i.e., X is separable.

(ix) Suppose $f \in Y^*$. By (i) there is an $F \in X^*$ such that $F = f$ on Y and $\|F\| = \|f\|$. Then $\phi(F) = F + Y^\perp$ is a well-defined element in X^*/Y^\perp . It is easy to prove that ϕ establishes an isometry between Y^* and X^*/Y^\perp .

As with the second isometry, we define $\psi(f)(x) = f(x + Y)$ for any $f \in (X/Y)^*$, and then find that $\psi : (X/Y)^* \rightarrow Y^\perp$ is an isometry and so $(X/Y)^*$ and Y^\perp are isometric.

(x) According to (vi), every $x \in X$ defines a unique element \tilde{x} in the second dual X^{**} of X and $\|\tilde{x}\| = \|x\|_X$, and consequently, $\tilde{\cdot}$ is always linear and an isometry from X into X^{**} . Therefore, X is not only identified with \tilde{X} but also regarded as a linear subspace of X^{**} . Since X is reflexive, every functional in X^{**} just arises in the fashion $\tilde{x}(L) = L(x)$ for $L \in X^*$ and $x \in X$; that is, the mapping $\tilde{\cdot}$ is surjective.

Suppose that Y is closed. Then (ix) tells us that $Y^* = X^*/Y^\perp$ and $(X^*/Y^\perp)^* = (Y^\perp)^\perp = \{f \in X^{**} : f(Y^\perp) = \{0\}\}$. Since X is viewed as a linear subspace of X^{**} , any element $y \in Y$ induces $f(y) = 0$ for all $f \in Y^\perp = \{f \in X^* : f(Y) = \{0\}\}$ and consequently, y is a member of $(Y^\perp)^\perp$. Meanwhile, if there is an element y_0 in $(Y^\perp)^\perp \setminus Y$, then the closedness of Y ensures $\inf_{y \in Y} \|y - y_0\|_X = \delta > 0$, and hence by (ii) there exists an $f \in X^*$ such that $f(y_0) = 1$, $f(y) = 0$ for all $y \in Y$ (this implies $f \in Y^\perp$), and $\|f\| = \delta^{-1}$. Note that X is reflexive. So $y_0 \in (Y^\perp)^\perp$ and $f \in Y^\perp$ yield $f(y_0) = 0$, contradicting $f(y_0) = 1$. Therefore, $(Y^\perp)^\perp = Y$, and consequently, $Y^{**} = Y$, as desired. \square

Problems

8.1. Let $p \in (1, \infty)$ and $q = p(p-1)^{-1}$. Prove Hölder's and Minkowski's inequalities of the discrete forms:

(i) For vectors $(x_j)_{j=1}^n$ and $(y_j)_{j=1}^n$ in \mathbf{R}^n , one has

$$\sum_{j=1}^n |x_j y_j| \leq \left(\sum_{j=1}^n |x_j|^p \right)^{\frac{1}{p}} \left(\sum_{j=1}^n |y_j|^q \right)^{\frac{1}{q}}$$

and

$$\left(\sum_{j=1}^n |x_j + y_j|^p \right)^{\frac{1}{p}} \leq \left(\sum_{j=1}^n |x_j|^p \right)^{\frac{1}{p}} + \left(\sum_{j=1}^n |y_j|^p \right)^{\frac{1}{p}}.$$

(ii) For real-valued sequences $(x_j)_{j=1}^\infty \in \ell_p$ and $(y_j)_{j=1}^\infty \in \ell_q$, one has

$$\sum_{j=1}^\infty |x_j y_j| \leq \left(\sum_{j=1}^\infty |x_j|^p \right)^{\frac{1}{p}} \left(\sum_{j=1}^\infty |y_j|^q \right)^{\frac{1}{q}}$$

and

$$\left(\sum_{j=1}^\infty |x_j + y_j|^p \right)^{\frac{1}{p}} \leq \left(\sum_{j=1}^\infty |x_j|^p \right)^{\frac{1}{p}} + \left(\sum_{j=1}^\infty |y_j|^p \right)^{\frac{1}{p}}.$$

8.2. Determine the dimensions of the following linear spaces.

(i) The set of vectors $x = (x_1, \dots, x_n)$ in \mathbf{R}^n with $\sum_{j=1}^n x_j = 0$;

(ii) The set $C[0, 1]$ of all continuous functions $f: [0, 1] \rightarrow \mathbf{R}$;

(iii) The set of all real-valued polynomials on $[0, 1]$.

8.3. Prove the following two results.

(i) On $C[0, 1]$ the sup-norm is not equivalent to any p -norm where $1 \leq p < \infty$;

(ii) If $C[0, 1]$ is equipped with the sup-norm, then the unit sphere of $C[0, 1]$ is not compact.

8.4. Prove that if X is a linear space over \mathbf{F} then any norm $\|\cdot\|: X \rightarrow \mathbf{R}$ is continuous on X , but also vector addition and scalar multiplication are continuous whenever $X \times Y$ is equipped with the norm $\|\cdot\|_X + \|\cdot\|_Y$.

8.5. Recall that ℓ_∞ is the space of all $x = (x_j)_{j=1}^\infty$ of real numbers x_j with $\|x\|_\infty = \sup_{j \in \mathbf{N}} |x_j| < \infty$. Prove that if ℓ_0 is the class of all elements $(x_j)_{j=1}^\infty \in \ell_\infty$ having only finitely many non-zero coordinates, then ℓ_0 is not closed in ℓ_∞ .

8.6. Prove that:

- (i) $\lim_{p \rightarrow \infty} \|f\|_{p, m_{id}, [0,1]} = \|f\|_{\infty, m_{id}, [0,1]}$ whenever $\|f\|_{\infty, m_{id}, [0,1]} < \infty$;
 (ii) $\int_{[0,1]} |f_1 f_2| dm_{id} \leq \|f_1\|_{1, m_{id}, [0,1]} \|f_2\|_{\infty, m_{id}, [0,1]}$ with equality if and only if $|f_2(x)| = \|f_2\|_{\infty, m_{id}, [0,1]}$ for m_{id} -a.e. $x \in [0, 1]$.

8.7. Prove that if $C[0, 1]$ is equipped with the sup-norm and $T : C[0, 1] \rightarrow \mathbf{R}$ is given by $T(f) = f(0)$, then T is bounded with $\|T\| = 1$.

8.8.

(i) Suppose that the real infinite matrix $[a_{kj}]$ satisfies $\sup_{k \in \mathbf{N}} \sum_{j=1}^{\infty} |a_{kj}| < \infty$. Define

$$T : x = (x_1, x_2, \dots) \rightarrow y = T(x) = \left(\sum_{j=1}^{\infty} a_{1j} x_j, \sum_{j=1}^{\infty} a_{2j} x_j, \dots \right).$$

Prove that $T : \ell_{\infty} \rightarrow \ell_{\infty}$ is bounded and $\|T\| = \sup_{k \in \mathbf{N}} \sum_{j=1}^{\infty} |a_{kj}|$.

(ii) Let $T : \ell_2 \rightarrow \ell_2$ be defined by

$$T(x) = (0, x_1, x_2, \dots) \text{ for } x = (x_1, x_2, \dots) \in \ell_2.$$

Prove $\|T(x)\|_2 = \|x\|_2$.

(iii) For a finite interval $[a, b]$ in \mathbf{R} , let $R^1[a, b]$ be the space of all real-valued Riemann integrable functions f on $[a, b]$ with $\|f\|_1 = \int_a^b |f(x)| dx < \infty$. Define $T(f)(x) = \int_a^x f(t) dt$. Prove that T is a bounded linear operator from $R^1[a, b]$ to itself with $\|T\| = b - a$.

8.9. Let $C(0, 1)$ be the space of all real-valued continuous functions on $(0, 1)$. Equip $C(0, 1)$ with the 2-norm $\|f\|_2 = \left(\int_0^1 |f(t)|^2 dt \right)^{1/2}$. Define

$$T : C(0, 1) \rightarrow C(0, 1) \text{ by } T(f)(t) = t f(t) \text{ for } t \in (0, 1).$$

Prove that T is bounded but not invertible.

8.10. Let $T : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be given by $T((x_1, x_2)) = (x_1, 0)$. Prove that T is linear, bounded, but not onto, and cannot map open sets to open sets in \mathbf{R}^2 .

8.11. For a normed linear space X over \mathbf{F} , prove the following results:

(i) If $T \in \mathcal{B}(X)$ and T^{-1} exists and belongs to $\mathcal{B}(X)$, then $(T^{-1})^n = (T^n)^{-1}$ holds for any $n \in \mathbf{N}$.

(ii) If $T, S \in \mathcal{B}(X)$ and TS has an inverse in $\mathcal{B}(X)$, then T and S must have inverses in $\mathcal{B}(X)$.

8.12. Let K be a convex set in a normed linear space $(X, \|\cdot\|_X)$ over \mathbf{R} ; that is,

$$x_1, x_2 \in K \text{ and } \lambda \in (0, 1) \Rightarrow \lambda x_1 + (1 - \lambda)x_2 \in K.$$

If 0 is an interior point of K , then the Minkowski functional p of K is defined by

$$p(x) = \inf\{\lambda > 0 : x \in \lambda K\} \text{ where } \lambda K = \{\lambda x : x \in K\}.$$

Prove the following results:

(i) The Minkowski functional of the unit ball in X is the norm $\|\cdot\|_X$.

(ii) The Minkowski functional p is positive homogeneous; that is,

$$p(\kappa x) = \kappa p(x) \text{ for } \kappa > 0 \text{ and } x \in X$$

and subadditive; that is,

$$p(x+y) \leq p(x) + p(y) \text{ for all } x, y \in X.$$

(iii) For each $x_0 \notin K$ there exists an $L \in \mathcal{B}(X, \mathbf{R})$ such that $L(x) \leq L(x_0)$ for $x \in K$.

8.13. Let X be a real normed linear space. By a hyperplane in X is meant a set of the form $H = \{x \in X : f(x) = r\}$ where f is a continuous linear functional on X and $r \in \mathbf{R}$. Prove the following separation theorem: If E_1 and E_2 are disjoint convex subsets of X and E_1 has an interior point, then there is a hyperplane which separates E_1 and E_2 ; that is, there exist a continuous functional $f \in \mathcal{B}(X, \mathbf{R})$ and an $r \in \mathbf{R}$ such that

$$E_1 \subseteq \{x \in X : f(x) \leq r\} \text{ and } E_2 \subseteq \{x \in X : f(x) \geq r\}.$$

8.14. Let X be a normed linear space over \mathbf{F} and $f : X \rightarrow \mathbf{F}$ be a linear functional. Prove that $f \in \mathcal{B}(X, \mathbf{F})$ if and only if $f^{-1}(\{0\})$ is closed in X .

8.15. Let X be a normed linear space over \mathbf{F} . Prove that if M is a closed linear subspace of X and $x \in X \setminus M$ then $M + \mathbf{F}x$ is closed.

8.16. If M is a finite-dimensional subspace of a normed linear space X over \mathbf{F} , prove that there is a closed linear subspace N such that $M \cap N = \{0\}$ and $M + N = X$.

8.17. *Krein's extension theorem.* Let X be a normed linear space over \mathbf{R} and $Y \subseteq X$ such that if $y_1, y_2 \in Y$ and $c \geq 0$ then $y_1 + y_2 \in Y$ and $cy_1 \in Y$. Define a partial order on X by declaring that $x_1 \preceq x_2$ if and only if $x_2 - x_1 \in Y$, and call a linear functional f on X Y -positive if $f(x) \geq 0$ for $x \in Y$. Let M be a subspace of X such that for each $x \in X$ there exists $y \in M$ with $x \preceq y$. Prove that if f is a $(Y \cap M)$ -positive linear functional on M , then there is a Y -positive linear functional F on X such that the restriction of F on M is the same as f .

Chapter 9

Banach Spaces via Operators and Functionals

Banach spaces, named after Stefan Banach who investigated them, are one of the central objects of study in linear functional analysis. Banach spaces are typically infinite-dimensional spaces containing functions. In this chapter we are concerned with the definition of a Banach space followed by some examples, some of the basic algebraic and topological properties of Banach spaces, continuous linear functionals, as well as bounded and compact linear operators over Banach spaces.

9.1 Definition and Beginning Examples

As before, \mathbf{F} always means the real field \mathbf{R} or the complex field \mathbf{C} .

Definition 9.1.1. *A normed linear space over \mathbf{F} which is complete in the metric generated by the norm is called a Banach space over \mathbf{F} .*

Clearly, if a linear space is a Banach space under one norm, then it is also a Banach space under any equivalent norm, and hence for each $n \in \mathbf{N}$, \mathbf{R}^n is complete under any norm. One more beginning example of Banach spaces is listed below.

Example 9.1.2.

(i) For a compact metric space (X, d_X) , suppose that $C(X, \mathbf{F})$ is equipped with the sup-norm $\|f\|_\infty = \sup_{x \in X} |f(x)|$ for $f \in C(X, \mathbf{F})$. Then it is a Banach space over \mathbf{F} , because of Theorem 7.3.7 (i).

(ii) For each $p \in [1, \infty]$, let $\ell_p(\mathbf{N}, \mathbf{F})$ be the space of all \mathbf{F} -valued sequences $\{x_j = (x_j^{(k)})_{k=1}^\infty\}_j$ satisfying

$$\|(x_j^{(k)})_{k=1}^\infty\|_p = \begin{cases} \left(\sum_{k=1}^\infty |x_j^{(k)}|^p\right)^{\frac{1}{p}}, & p \in [1, \infty) \\ \sup_{k \in \mathbf{N}} |x_j^{(k)}|, & p = \infty \end{cases}.$$

Then $(\ell_p(\mathbf{N}, \mathbf{F}), \|\cdot\|_p)$ is a Banach space over \mathbf{F} . To see this, assume that $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence in $\ell_p(\mathbf{N}, \mathbf{F})$, and write $x_j = (x_j^{(1)}, x_j^{(2)}, \dots)$. Because of $\|\cdot\|_p \geq \|\cdot\|_\infty$, for any $\varepsilon > 0$ we may find $N \in \mathbf{N}$ such that $m, n > N \Rightarrow \|x_n - x_m\|_p < \varepsilon$ which in turn implies that $\|x_n - x_m\|_\infty < \varepsilon$, so for each k , $|x_n^{(k)} - x_m^{(k)}| < \varepsilon$. In other words, if $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence in $\ell_p(\mathbf{N}, \mathbf{F})$ then $\{x_j^{(k)}\}_{j=1}^\infty$ is a Cauchy sequence in \mathbf{F} . Since \mathbf{F} is complete, for each k we have that $|x_j^{(k)} - y^{(k)}| \rightarrow 0$ as $j \rightarrow \infty$. Note that this does not imply by itself $x_j \rightarrow y = (y^{(1)}, y^{(2)}, \dots)$. However, if we know that $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence, then it does. In fact, we prove this for $p < \infty$ but the $p = \infty$ case is similar. Fix $\varepsilon > 0$, and use the Cauchy criterion to find an $n_0 \in \mathbf{N}$ such that

$$n, m > n_0 \Rightarrow \sum_{k=1}^\infty |x_n^{(k)} - x_m^{(k)}|^p < \varepsilon.$$

Now fix $n \in \mathbf{N}$ and let $m \rightarrow \infty$ to see $\sum_{k=1}^\infty |x_n^{(k)} - y^{(k)}|^p \leq \varepsilon$. This inequality means

$$\|x_n - y\|_p \leq \varepsilon^{\frac{1}{p}} \text{ and } y = (y^{(1)}, y^{(2)}, \dots) \in \ell_p(\mathbf{N}, \mathbf{F}).$$

To give another important class of Banach spaces, we need one more definition and theorem.

Definition 9.1.3. For a sequence $\{x_j\}_{j=1}^\infty$ in a normed linear space $(X, \|\cdot\|_X)$ over \mathbf{F} , we say that the series $\sum_{j=1}^\infty x_j$ is absolutely convergent provided that $\sum_{j=1}^\infty \|x_j\|_X < \infty$. Moreover, a series $\sum_{j=1}^\infty x_j$ is said to converge to x in X provided that $\lim_{n \rightarrow \infty} \|\sum_{j=1}^n x_j - x\|_X = 0$.

Theorem 9.1.4. *Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} . Then it is a Banach space if and only if the absolute convergence of $\sum_{j=1}^{\infty} x_j$ implies its convergence.*

Proof. On the one hand, suppose that $(X, \|\cdot\|_X)$ is a Banach space. Consider the sequence of partial sums $s_k = \sum_{j=1}^k x_j$. Since $\sum_{j=1}^{\infty} x_j$ is absolutely convergent, we conclude that

$$\|s_m - s_k\|_X \leq \sum_{j=k+1}^m \|x_j\|_X \rightarrow 0 \text{ as } m > k \rightarrow \infty.$$

It follows that $\{s_m\}_{m=1}^{\infty}$ is a Cauchy sequence, and consequently this sequence converges since X is complete. Thus, $\sum_{j=1}^{\infty} x_j$ converges.

On the other hand, assume $\{x_j\}_{j=1}^{\infty}$ is a Cauchy sequence in $(X, \|\cdot\|_X)$. Then for each $k \in \mathbf{N}$ there is an $n_k \in \mathbf{N}$ such that

$$n, j \geq n_k \Rightarrow \|x_n - x_j\|_X < 2^{-k}.$$

Without loss of generality, we may assume $n_{k+1} \geq n_k$. This yields that $\{x_{n_k}\}_{k=1}^{\infty}$ is a subsequence of $\{x_k\}_{k=1}^{\infty}$. Set $y_1 = x_{n_1}$ and $y_k = x_{n_k} - x_{n_{k-1}}$ when $k \geq 2$. Note that

$$\sum_{k=1}^l \|y_k\|_X < \|y_1\|_X + \sum_{k=2}^l 2^{1-k} \leq \|y_1\|_X + 1.$$

So $\{\sum_{k=1}^l \|y_k\|_X\}_{l=1}^{\infty}$ is increasing and bounded, and hence is convergent. By hypothesis, $\sum_{k=1}^{\infty} y_k$ is convergent. Since $\sum_{k=1}^l y_k = x_{n_l}$, it follows that $\{x_{n_l}\}_{l=1}^{\infty}$ is convergent in $(X, \|\cdot\|_X)$. This, together with the fact that $\{x_k\}_{k=1}^{\infty}$ is a Cauchy sequence, infers that $\{x_k\}_{k=1}^{\infty}$ is convergent. Therefore, $(X, \|\cdot\|_X)$ is a Banach space. \square

Remark 9.1.5. *It is worthwhile to point out that Theorem 9.1.4 is clearly not true for general normed linear spaces. For example, for $j \in \mathbf{N}$ and $p \in [1, \infty)$ let*

$$h_j(x) = \begin{cases} 0 & , x \in [-1, -j^{-1}] \\ 1 + jx & , x \in [-j^{-1}, 0] \\ 1 & , x \in [0, 1] \end{cases} \quad \text{and } f_j = h_{j+1} - h_j.$$

Then

$$f_j \in C([-1, 1], \mathbf{R}) \text{ and } \sum_{j=1}^{\infty} \|f_j\|_{p, m_{id}, [-1, 1]} = \sum_{j=1}^{\infty} (j+1)^{-1} (j(p+1))^{-\frac{1}{p}} < \infty,$$

but $\sum_{j=1}^{\infty} f_j = \lim_{n \rightarrow \infty} (h_{n+1} - h_1)$ with respect to $\|\cdot\|_{p, m_{id}, [-1, 1]}$ is outside $C([-1, 1], \mathbf{R})$ since $\lim_{n \rightarrow \infty} \|h_{n+1} - 1_{[0, 1]}\|_{p, m_{id}, [-1, 1]} = 0$. This example indicates that $C([-1, 1], \mathbf{R})$ is not complete under $\|\cdot\|_{p, m_{id}, [-1, 1]}$. Actually, its completion is $LRS_{id}^p([-1, 1], \mathbf{R})$.

Example 9.1.6. Let $LRS_g^p(E, \mathbf{F})$ be as in Example 8.1.14. Then each $LRS_g^p(E, \mathbf{F})$ is a Banach space whenever $p \in [1, \infty]$. Actually, in the case $p \in [1, \infty)$, the result can be demonstrated via Theorem 9.1.4. As with $p = \infty$, suppose that $\{f_j\}_{j=1}^{\infty}$ is a Cauchy sequence in $LRS_g^{\infty}(E, \mathbf{F})$. Clearly, we may assume $|f_j(x)| \leq \|f_k\|_{\infty, m_g, E}$ for all $x \in E$ by replacing f_j with an equivalent function for which this is true. Then $\{f_j(x)\}_{j=1}^{\infty}$ is a Cauchy sequence in \mathbf{F} for each $x \in E$ and we can thus get an \mathbf{F} -valued function f on E such that $\lim_{k \rightarrow \infty} f_k(x) = f(x)$ for each $x \in E$. This in turns yields $\lim_{k \rightarrow \infty} \|f - f_k\|_{\infty, m_g, E} = 0$.

Interestingly, Example 9.1.6 leads to a criterion for a quotient space to be complete.

Theorem 9.1.7. Let $(X, \|\cdot\|_X)$ be Banach space over \mathbf{F} and $Y \subseteq X$. Then X/Y is a Banach space under the norm $\|x + Y\|_{X/Y} = \inf_{z \in x+Y} \|z\|_X$ if and only if Y is closed.

Proof. If Y is not closed, then Theorem 8.1.13 tells us that $\|x + Y\|_{X/Y}$ is not a norm. So, it remains to prove that if Y is closed then X/Y is complete. In doing so, suppose that $\sum_{k=1}^{\infty} (x_k + Y)$ is an absolutely convergent series in X/Y . For each $k \in \mathbf{N}$ choose $z_k \in x_k + Y$ such that

$$\|z_k\|_X \leq \|x_k + Y\|_{X/Y} + 2^{-k}.$$

Since X is a Banach space, it follows that $\sum_{k=1}^{\infty} z_k$ is absolutely convergent and thus convergent to some $x \in X$ thanks to Theorem 9.1.4. Note that

$$\left\| x + Y - \sum_{k=1}^n (x_k + Y) \right\|_{X/Y} \leq \left\| x - \sum_{k=1}^n z_k \right\|_X \rightarrow 0 \text{ as } n \rightarrow \infty.$$

So we obtain that $\sum_{k=1}^{\infty} (x_k + Y)$ converges to $x + Y \in X/Y$. Namely, X/Y is complete due to Theorem 9.1.4. \square

In many situations it makes sense to multiply elements of a Banach space together.

Definition 9.1.8. Let $(X, \|\cdot\|_X)$ be a Banach space over \mathbf{F} . If there is a multiplication $(x, y) \mapsto xy$ from $X \times X \rightarrow X$ such that for any $x, y, z \in X$ and $\alpha \in \mathbf{F}$,

- (i) $x(yz) = (xy)z$;
- (ii) $x(y+z) = xy + xz$;
- (iii) $(x+y)z = xz + yz$;
- (iv) $\alpha(xy) = (\alpha x)y = x(\alpha y)$;
- (v) $\|xy\|_X \leq \|x\|_X \|y\|_X$,

then X is called a Banach algebra.

Example 9.1.9.

- (i) $(C[0, 1], \|\cdot\|_\infty)$ is a Banach algebra with $(fg)(x) = f(x)g(x)$.
- (ii) $LRS_{id}^1(\mathbf{R}, \mathbf{F})$ is a Banach algebra under the convolution

$$f_1 * f_2(x) = \int_{\mathbf{R}} f_1(x-t)f_2(t)dm_{id}(t).$$

Here, we should note

$$f \in LRS_{id}^1(\mathbf{R}, \mathbf{F}) \Rightarrow \int_{\mathbf{R}} f dm_{id} = \int_{\mathbf{R}} \Re f dm_{id} + i \int_{\mathbf{R}} \Im f dm_{id}.$$

(iii) If $X = \mathbf{R}^n$, $n \in \mathbf{N}$, then by choosing a basis for \mathbf{R}^n we may identify $\mathcal{B}(\mathbf{R}^n)$ with the space of $n \times n$ real matrices which is a Banach algebra under the usual matrix multiplication.

(iv) If $(X, \|\cdot\|_X)$ is a Banach space over \mathbf{F} , then $(\mathcal{B}(X), \|\cdot\|)$ is a Banach algebra under the composition $ST = S \circ T$.

Beyond Example 9.1.9 (iv), we can derive more information on all bounded linear operators on a given Banach space.

Theorem 9.1.10. Let $(X, \|\cdot\|_X)$ be a Banach space over \mathbf{F} .

(i) If $T \in \mathcal{B}(X)$ satisfies $\|T\| < 1$ and I is the identity element in $\mathcal{B}(X)$, then $I - T \in \mathcal{B}(X)$ is invertible and

$$(I - T)^{-1} = \lim_{n \rightarrow \infty} (I + T + T^2 + \cdots + T^n) \text{ in } (\mathcal{B}(X), \|\cdot\|),$$

where T^k is the k -th composition $\underbrace{T \circ \cdots \circ T}_k$ for $k \in \mathbf{N}$.

(ii) If I stands for the class of all invertible operators in $\mathcal{B}(X)$, then I is an open set in $\mathcal{B}(X)$.

Proof. (i) Fix $x \in X$. Since $m, n \in \mathbf{N}$ and $m > n$ imply

$$\begin{aligned}
 & \| (I + T + T^2 + \cdots + T^m)(x) - (I + T + T^2 + \cdots + T^n)(x) \|_X \\
 &= \| T^{n+1}(x) + \cdots + T^m(x) \|_X \\
 &\leq \| T^{n+1}(x) \|_X + \cdots + \| T^m(x) \|_X \\
 &\leq (\| T^{n+1} \| + \cdots + \| T^m \|) \| x \|_X \\
 &\leq \left(\sum_{j=n+1}^{\infty} \| T \|^j \right) \| x \|_X \\
 &= \| T \|^{n+1} (1 - \| T \|)^{-1} \| x \|_X \\
 &\rightarrow 0 \text{ as } n \rightarrow \infty,
 \end{aligned}$$

we conclude that $\{ (I + T + T^2 + \cdots + T^n)(x) \}_{n=1}^{\infty}$ is a Cauchy sequence in X . Note that $(X, \| \cdot \|_X)$ is a Banach space. So the sequence converges to a limit $y \in X$. Let $y = A(x)$. It is not hard to show that $A : X \rightarrow X$ is a linear operator on X . Furthermore, letting $m \rightarrow \infty$, we have

$$\| A(x) - (I + T + T^2 + \cdots + T^n)(x) \|_X \leq \| T \|^{n+1} (1 - \| T \|)^{-1} \| x \|_X,$$

so that $A - (I + T + T^2 + \cdots + T^n) \in \mathcal{B}(X)$, and thus $A \in \mathcal{B}(X)$. Because

$$\| A - (I + T + T^2 + \cdots + T^n) \| \leq \| T \|^{n+1} (1 - \| T \|)^{-1} \rightarrow 0 \text{ as } n \rightarrow \infty,$$

we derive that $I + T + T^2 + \cdots + T^n \rightarrow A$ as $n \rightarrow \infty$. It remains to verify $A = (I - T)^{-1}$. For any $x \in X$ we have

$$\begin{aligned}
 ((I - T)A)(x) &= ((I - T) \lim_{n \rightarrow \infty} (I + T + T^2 + \cdots + T^n))(x) \\
 &= (I - T) \left(\lim_{n \rightarrow \infty} (I(x) + T(x) + T^2(x) + \cdots + T^n(x)) \right) \\
 &= \lim_{n \rightarrow \infty} (x - T^{n+1}(x)).
 \end{aligned}$$

But

$$\| T^{n+1}(x) \|_X \leq \| T \|^{n+1} \| x \|_X \rightarrow 0 \text{ as } n \rightarrow \infty,$$

so that $T^{n+1}(x) \rightarrow 0$ as $n \rightarrow \infty$, and consequently $((I - T)A)(x) = x$. Similarly, we have $(A(I - T))(x) = x$, whence getting $A = (I - T)^{-1}$.

(ii) Note that $\|T^{-1}\| \neq 0$ for $T \in I$. So, to prove that the open ball

$$\{S \in \mathcal{B}(X) : \|T - S\| < \|T^{-1}\|^{-1}\}$$

is a subset of I , we just show that every element S of this ball is invertible. Using

$$\|(T - S)T^{-1}\| \leq \|T - S\| \|T^{-1}\| < 1,$$

we obtain that $ST^{-1} = I - (T - S)T^{-1}$ is invertible and thus $S = (ST^{-1})T$ is invertible. \square

Remark 9.1.11. When $(X, \|\cdot\|_X)$ is a Banach space over \mathbf{F} and T belongs to $\mathcal{B}(X)$, we can define an operator

$$e^T = I + T + \frac{T^2}{2!} + \frac{T^3}{3!} + \cdots,$$

which makes sense since

$$\|e^T\| \leq 1 + \|T\| + \frac{\|T\|^2}{2!} + \frac{\|T\|^3}{3!} + \cdots = \exp \|T\|.$$

With this notion, we have that if $x(t) = (x_1(t), x_2(t), \dots, x_n(t)) \in \mathbf{R}^n$, then the system

$$\frac{dx}{dt} = \left(\frac{dx_1(t)}{dt}, \frac{dx_2(t)}{dt}, \dots, \frac{dx_n(t)}{dt} \right) = x(t)A \text{ subject to } x(0) = a \in \mathbf{R}^n,$$

where A is an $n \times n$ matrix, has a solution $x(t) = ae^{tA}$.

9.2 Uniform Boundedness - Open Map - Closed Graph

In this section, we proceed to discuss three fundamental results – the uniform boundedness principle (sometimes known as the Banach-Steinhaus theorem due to Stefan Banach and Hugo Dyonizy Steinhaus), open mapping theorem and closed graph theorem.

Below is the so-called uniform boundedness principle.

Theorem 9.2.1. Let $(X, \|\cdot\|_X)$ be a Banach space and $(Y, \|\cdot\|_Y)$ a normed linear space over \mathbf{F} . Let $\{T_\alpha\}_{\alpha \in I}$ be a family of bounded linear operators $T_\alpha : X \rightarrow Y$ indexed by the nonempty set I . If $\sup_{\alpha \in I} \|T_\alpha(x)\|_Y < \infty$ for each $x \in X$, then $\sup_{\alpha \in I} \|T_\alpha\| < \infty$.

Proof. For each $j \in \mathbf{N}$ let

$$X_j = \{x \in X : \sup_{\alpha \in I} \|\mathsf{T}_\alpha(x)\|_Y \leq j\} = \cap_{\alpha \in I} \{x \in X : \|\mathsf{T}_\alpha(x)\|_Y \leq j\}.$$

Then each X_j is a closed subset of X and $X = \cup_{j=1}^\infty X_j$. Since X is complete, X is of second category. By the Baire category theorem (i.e., Theorem 6.3.17) we see that $X_j^\circ \neq \emptyset$ for some $j = j_0 \in \mathbf{N}$ and consequently, there is an $x_0 \in X$ and an $r \in (0, \infty)$ such that the ball $B_r(x_0) \subseteq X_{j_0}$. This means that for $x \in B_r(x_0)$ we have $\sup_{\alpha \in I} \|\mathsf{T}_\alpha(x)\|_Y \leq j_0$. Consequently, for any $x \in B_r(0)$ - the origin-centered ball we obtain $x_0 - x \in B_r(x_0)$ and so for any $\alpha \in I$,

$$\|\mathsf{T}_\alpha(x)\|_Y \leq \|\mathsf{T}_\alpha(x - x_0)\|_Y + \|\mathsf{T}_\alpha(x_0)\|_Y \leq 2j_0.$$

Thus, for each $x \in \overline{B_1(0)}$ we get $\|\mathsf{T}_\alpha(x)\|_Y \leq 4j_0r^{-1}$. This yields $\sup_{\alpha \in I} \|\mathsf{T}_\alpha\| \leq 4j_0r^{-1}$, as required. \square

The boundedness of a sequence of linear operators is useful in the study of convergence of operator sequences. To see this, we introduce two types of convergence involving operator sequences.

Definition 9.2.2. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces over \mathbf{F} .

(i) A sequence $\{\mathsf{T}_n\}_{n=1}^\infty$ in $\mathcal{B}(X, Y)$ is said to be uniformly convergent provided that there is a $\mathsf{T} \in \mathcal{B}(X, Y)$ such that $\lim_{n \rightarrow \infty} \|\mathsf{T}_n - \mathsf{T}\| = 0$.

(ii) A sequence $\{\mathsf{T}_n\}_{n=1}^\infty$ in $\mathcal{B}(X, Y)$ is said to be strongly (or pointwise) convergent on X provided that the sequence $\{\mathsf{T}_n(x)\}_{n=1}^\infty$ is convergent in Y for any $x \in X$. Moreover, if there is a $\mathsf{T} \in \mathcal{B}(X, Y)$ such that $\lim_{n \rightarrow \infty} \|\mathsf{T}_n(x) - \mathsf{T}(x)\|_Y = 0$ for all $x \in X$, then $\{\mathsf{T}_n\}_{n=1}^\infty$ is said to be strongly convergent to T .

Example 9.2.3. Clearly, the uniform convergence implies the strong convergence, but not conversely. Consider $\ell_p(\mathbf{N}, \mathbf{F})$, $p \in [1, \infty)$. For each $n \in \mathbf{N}$ define

$$\mathsf{T}_n(x) = (x_n, x_{n+1}, \dots) \text{ for all } x = (x_1, x_2, \dots) \in \ell_p(\mathbf{N}, \mathbf{F}).$$

Then T_n is in $\mathcal{B}(\ell_p(\mathbf{N}, \mathbf{F}))$ with $\|\mathsf{T}_n\| \leq 1$. Note that if $x = (x_1, x_2, \dots) \in \ell_p(\mathbf{N}, \mathbf{F})$ then

$$\lim_{n \rightarrow \infty} \|\mathsf{T}_n(x)\|_p = \lim_{n \rightarrow \infty} \left(\sum_{j=n}^\infty |x_j|^p \right)^{\frac{1}{p}} = 0.$$

So $\{T_n\}_{n=1}^\infty$ is strongly convergent to 0. Meanwhile, for $n \in \mathbf{N}$ let $e_n = (\underbrace{0, \dots, 0}_{n-1}, 1, 0, \dots)$. Then $\|e_n\|_p = 1$ and $T_n(e_n) = (1, 0, 0, \dots)$ and hence $\|T_n\| \geq \|T_n(e_n)\|_p = 1$. This shows that $\{T_n\}_{n=1}^\infty$ is not uniformly convergent to 0.

Here is a special consequence of the uniform boundedness principle.

Theorem 9.2.4. *Let $(X, \|\cdot\|_X)$ be a Banach space over \mathbf{F} .*

(i) *Suppose that $(Y, \|\cdot\|_Y)$ is a normed linear space over \mathbf{F} . If a sequence $\{T_n\}_{n=1}^\infty$ in $\mathcal{B}(X, Y)$ is strongly convergent, then there exists a $T \in \mathcal{B}(X, Y)$ such that $\{T_n\}_{n=1}^\infty$ is strongly convergent to T .*

(ii) *Suppose that $(Y, \|\cdot\|_Y)$ is a Banach space over \mathbf{F} . Then a sequence $\{T_n\}_{n=1}^\infty$ in $\mathcal{B}(X, Y)$ is strongly convergent if and only if $\sup_{n \in \mathbf{N}} \|T_n\| < \infty$ and $\{T_n(x)\}_{n=1}^\infty$ is convergent in Y for any x in a dense subset D of X .*

Proof. (i) Let $\{T_n\}_{n=1}^\infty$ in $\mathcal{B}(X, Y)$ be strongly convergent. Then for each $x \in X$ the sequence $\{T_n(x)\}_{n=1}^\infty$ is convergent in Y and hence defines a linear operator T on X by setting $T(x) = \lim_{n \rightarrow \infty} T_n(x)$. The key is to show that $T \in \mathcal{B}(X, Y)$. Note that $\{T_n(x)\}_{n=1}^\infty$ is bounded in Y . So, from the uniform boundedness principle it turns out that there is a constant $C > 0$ such that $\sup_{n \in \mathbf{N}} \|T_n\| \leq C$. Hence

$$\|T_n(x)\|_Y \leq C\|x\|_X \quad \text{for all } x \in X.$$

This implies $\|T(x)\|_Y \leq C\|x\|_X$ for all $x \in X$, showing that T is bounded. The definition of T means that $\{T_n\}_{n=1}^\infty$ converges strongly to T .

(ii) The necessity follows from the uniform boundedness principle. As for the sufficiency, we assume the statement after the if and only if. Given $\varepsilon > 0$. Then for any $x \in X$ there is an element $z \in D$ such that $\|x - z\|_X < \varepsilon$, and there is an $N \in \mathbf{N}$ such that

$$m > n > N \Rightarrow \|T_m(z) - T_n(z)\|_Y < \varepsilon.$$

Consequently,

$$\begin{aligned} & \|T_m(x) - T_n(x)\|_Y \\ & \leq \|T_m(x) - T_m(z)\|_Y + \|T_m(z) - T_n(z)\|_Y + \|T_n(z) - T_n(x)\|_Y \\ & \leq \|T_m\| \|x - z\|_X + \varepsilon + \|T_n\| \|x - z\|_X \\ & \leq (1 + 2 \sup_{n \in \mathbf{N}} \|T_n\|) \varepsilon. \end{aligned}$$

That is to say, $\{T_n(x)\}_{n=1}^\infty$ is a Cauchy sequence in Y and thus convergent thanks to the completeness of $(Y, \|\cdot\|_Y)$. \square

Recall that a continuous mapping between two normed linear spaces has the property that the pre-image of any open set is open, but in general the image of an open set is not open. But, continuous linear operators between Banach spaces cannot do this. This is the content of the open mapping theorem as follows.

Theorem 9.2.5. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces over \mathbf{F} . If T is a continuous linear operator from X onto Y , then T is an open map, namely, $T(O)$ is open in Y whenever O is open in X .*

Proof. We split the argument into three steps. In what follows, we employ $B_r^X(x)$ and $B_r^Y(y)$ to denote the balls of radius $r > 0$ centered at $x \in X$ and $y \in Y$ respectively. In particular, we write $B_r^X = B_r^X(0)$ and $B_r^Y = B_r^Y(0)$ respectively.

Step 1. We prove that for any $\varepsilon > 0$ there is a $\delta > 0$ such that $B_\delta^Y \subseteq T(\overline{B_{2\varepsilon}^X})$. To see this, note that

$$X = \bigcup_{n=1}^\infty nB_\varepsilon^X = \bigcup_{n=1}^\infty \{nx : x \in B_\varepsilon^X\},$$

and $T \in \mathcal{B}(X, Y)$ is surjective. So we have

$$Y = T(X) = \bigcup_{n=1}^\infty nT(B_\varepsilon^X) = \bigcup_{n=1}^\infty \overline{nT(B_\varepsilon^X)}.$$

Since $(Y, \|\cdot\|_Y)$ is a Banach space, we conclude that from the Baire category theorem (i.e. Theorem 6.3.17) that some $\overline{nT(B_\varepsilon^X)}$ is not nowhere dense in Y - namely $-\left(\overline{nT(B_\varepsilon^X)}\right)^\circ \neq \emptyset$, and so that

$$B_r^Y(z) = \{y \in Y : \|y - z\|_Y < r\} \subseteq \overline{nT(B_\varepsilon^X)} = \{ny : y \in \overline{T(B_\varepsilon^X)}\}$$

for some $z \in Y$ and some $r > 0$. Thus $\overline{T(B_\varepsilon^X)}$ must contain the ball $B_\delta^Y(y_0)$ where $y_0 = n^{-1}z$ and $\delta = n^{-1}r$. It follows that

$$V = \{y_1 - y_2 : y_1, y_2 \in B_\delta^Y(y_0)\} \subseteq T(\overline{U}) \quad \text{where } U = \{x_1 - x_2 : x_1, x_2 \in B_\varepsilon^X\} \subseteq B_{2\varepsilon}^X.$$

Thus, $\overline{T(B_{2\varepsilon}^X)} \supseteq V$. Any point $y \in B_\delta^Y$ can be written as $y = (y + y_0) - y_0$, so $B_\delta^Y \subseteq V$, as desired.

Step 2. We further prove that for any $\varepsilon > 0$ there is a $\delta > 0$ such that $B_\delta^Y \subseteq T(B_{2\varepsilon}^X)$.

To do so, choose $\{\varepsilon_n\}_{n=1}^\infty$ with $\varepsilon_n > 0$ and $\sum_{n=1}^\infty \varepsilon_n < \varepsilon$. By *Step 1* there is a sequence of positive numbers $\{\delta_n\}_{n=1}^\infty$ such that $B_{\delta_n}^Y \subseteq T(\overline{B_{2\varepsilon_n}^X})$. Without loss of generality, we may assume $\lim_{n \rightarrow \infty} \delta_n = 0$.

Let $y \in B_{\delta_1}^Y$. Then $y \in T(\overline{B_{2\varepsilon_1}^X})$ and hence there is a point $x_1 \in B_{2\varepsilon_1}^X$ such that $\|y - T(x_1)\|_Y < \delta_2$. Since $y - T(x_1) \in B_{\delta_2}^Y$, we conclude that there is a point $x_2 \in B_{2\varepsilon_2}^X$ such that

$$\|y - T(x_1) - T(x_2)\|_Y < \delta_3.$$

Continuing this process, we obtain a sequence $\{x_n\}_{n=1}^\infty$ such that $x_n \in B_{2\varepsilon_n}^X$ and

$$\left\| y - T\left(\sum_{k=1}^n x_k\right) \right\|_Y < \delta_{n+1}.$$

Since $\|x_n\|_X < 2\varepsilon_n$, we conclude that $\sum_{n=1}^\infty x_n$ is absolutely convergent and hence by Theorem 9.1.4 that $x = \sum_{n=1}^\infty x_n$ makes sense in $(X, \|\cdot\|_X)$. This implies

$$\|x\|_X \leq \sum_{n=1}^\infty \|x_n\|_X < 2 \sum_{n=1}^\infty \varepsilon_n < 2\varepsilon.$$

Since the map T is continuous and $\delta_n \rightarrow 0$, we get $y = T(x)$. In other words, for any $y \in B_\delta^Y$ where $\delta = \delta_1$, we have found a point $x \in B_{2\varepsilon}^X$ such that $T(x) = y$, whence implying the desired inclusion.

Step 3. We prove that if O is open in X then for any point $x \in O$, there exists a $\delta > 0$ such that $B_\delta^Y(T(x)) \subseteq T(O)$.

In fact, if $x \in O$, then there exists an $\varepsilon > 0$ such that $B_{2\varepsilon}^X(x) \subseteq O$. By *Step 2*, we have

$$B_\delta^Y \subseteq T(B_{2\varepsilon}^X) \text{ for some } \delta > 0.$$

Hence

$$B_\delta^Y(T(x)) = T(x) + B_\delta^Y \subseteq T(x) + T(B_{2\varepsilon}^X) = T(x + B_{2\varepsilon}^X) = T(B_{2\varepsilon}^X(x)) \subseteq T(O).$$

Of course, this step completes the proof of the theorem. \square

Remark 9.2.6. *The open mapping theorem cannot ensure that the image of a closed linear subspace of the domain is closed. For example, let $T : \ell_2(\mathbf{N}, \mathbf{R}) \mapsto \ell_2(\mathbf{N}, \mathbf{R})$ be given by $(T(x))_j = x_{2j}$ for $x = \{x_j\}_{j=1}^\infty$, and set S be a linear subspace of $\ell_2(\mathbf{N}, \mathbf{R})$ comprising all vectors x with $x_{2j} = x_{2j-1}(2j)^{-1}$, $j \in \mathbf{N}$. Note that S is the intersection of the kernels $\{x \in \ell_2(\mathbf{N}, \mathbf{R}) : L_k(x) = 0\}$ of a countable*

collection of bounded linear functionals $\{L_k\}$ on $\ell_2(\mathbf{N}, \mathbf{R})$ (cf. Theorem 9.3.2). So, C is closed. But, $\mathcal{T}(C)$ is not closed since $\ell_2(\mathbf{N}, \mathbf{R}) \neq \mathcal{T}(S)$ and $\mathcal{T}(S)$ contains all points with finitely many non-zero terms - this means $\overline{\mathcal{T}(S)} = \ell_2(\mathbf{N}, \mathbf{R})$

As an application of the open mapping theorem we establish a general property of inverse mappings.

Theorem 9.2.7. *Let X and Y be Banach spaces over \mathbf{F} and let $\mathcal{T} \in \mathcal{B}(X, Y)$. If \mathcal{T} is bijective, then the inverse operator \mathcal{T}^{-1} of \mathcal{T} is a bounded linear map.*

Proof. It suffices to prove the continuity of \mathcal{T}^{-1} . Since $\mathcal{T} = (\mathcal{T}^{-1})^{-1}$ maps the Banach space X onto the Banach space Y , we conclude from Theorem 9.2.5 that \mathcal{T} maps open sets in X to open sets in Y . This amounts to saying that \mathcal{T}^{-1} is continuous. \square

Corollary 9.2.8. *Let $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ be two norms defined on a Banach space X over \mathbf{F} . If there is a constant $C_1 > 0$ such that $\|x\|_{(1)} \leq C_1 \|x\|_{(2)}$ for all $x \in X$, then there exists another constant $C_2 > 0$ such that $\|x\|_{(2)} \leq C_2 \|x\|_{(1)}$ for all $x \in X$. Consequently, both norms are equivalent.*

Proof. Consider the identity operator

$$I : (X, \|\cdot\|_{(2)}) \rightarrow (X, \|\cdot\|_{(1)}) \text{ where } I(x) = x \text{ for all } x \in X.$$

Clearly, I is bounded. By Theorem 9.2.7 we see that I^{-1} is also bounded, whence getting the norm inequality in the other direction. \square

Definition 9.2.9. *Given two normed linear spaces X and Y over \mathbf{F} , let $\mathcal{T} : X \rightarrow Y$ be a linear operator. Then the graph of \mathcal{T} is defined as*

$$G(\mathcal{T}) = \{(x, y) \in X \times Y : y = \mathcal{T}(x)\}.$$

Moreover, \mathcal{T} is said to be closed when $G(\mathcal{T})$ is a closed subset of $X \times Y$.

The following result is called the closed graph theorem.

Theorem 9.2.10. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces over \mathbf{F} . Then a linear operator $\mathcal{T} : X \rightarrow Y$ is bounded if and only if \mathcal{T} is closed.*

Proof. We initially observe that $X \times Y$ is a Banach space under the norm $\|(x, y)\| = \|x\|_X + \|y\|_Y$. Addition and scalar multiplication are defined in the expected manner:

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2) \text{ and } \alpha(x, y) = (\alpha x, \alpha y).$$

The completeness of $(X \times Y, \|\cdot\|)$ follows readily from the completeness of $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$.

On the one hand, suppose that $T : X \rightarrow Y$ is bounded. To prove that T is closed, i.e., $G(T)$ is closed, we assume

$$\lim_{n \rightarrow \infty} \|x_n - x\|_X = 0 \text{ and } \lim_{n \rightarrow \infty} \|T(x_n) - y\|_Y = 0.$$

The boundedness of T implies $\lim_{n \rightarrow \infty} \|T(x_n) - T(x)\|_Y = 0$. These two limits plus the triangle inequality imply $T(x) = y$. This means $(x, y) \in G(T)$. Thus T is closed.

On the other hand, suppose that T is closed. To verify $T \in \mathcal{B}(X, Y)$, we consider the projection map P from $G(T)$ onto X via: $P(x, T(x)) = x$. Clearly, P is linear, bijective and bounded. From Theorem 9.2.7 it turns out that P^{-1} is a bounded linear map from X onto $G(T)$, so there is a constant $C > 0$ such that

$$\|x\|_X + \|T(x)\|_Y = \|(x, T(x))\| = \|P^{-1}(x)\| \leq C\|x\|_X \text{ for all } x \in X.$$

Consequently, T is bounded. □

Remark 9.2.11. *It is worth mentioning that the second part of the argument for Theorem 9.2.10 has used the assumption that $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are complete. Whenever this assumption is removed, the conclusion fails, i.e., there exist a linear operator T_0 and a non-Banach space $(X_0, \|\cdot\|_{X_0})$ such that T_0 is closed but unbounded on $(X_0, \|\cdot\|_{X_0})$; see also Problem 9.8.*

9.3 Dual Banach Spaces by Examples

When dealing with a normed linear space X over the base field \mathbf{F} , one typically is only interested in the continuous linear functionals from the space into \mathbf{F} . According to Theorem 8.3.7 and Definition 8.4.1, these form a normed linear space. In this section, we will not only see that this new normed linear space is indeed a Banach space, but also completely describe the dual spaces of three typical Banach spaces.

We begin with a more general result.

Theorem 9.3.1. *Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} . If $(Y, \|\cdot\|_Y)$ is a Banach space over \mathbf{F} , then $\mathcal{B}(X, Y)$ is a Banach space over \mathbf{F} .*

Proof. If $\{\mathsf{T}_j\}_{j=1}^\infty$ is a Cauchy sequence in $\mathcal{B}(X, Y)$, then it is bounded and so there is a constant $C > 0$ such that $\|\mathsf{T}_j(x)\|_Y \leq C\|x\|_X$ for all $x \in X$ and $j \in \mathbf{N}$. Since

$$\|\mathsf{T}_j(x) - \mathsf{T}_k(x)\|_Y \leq \|\mathsf{T}_j - \mathsf{T}_k\| \|x\|_X \rightarrow 0 \text{ as } j \geq k \rightarrow \infty,$$

$\{\mathsf{T}_j(x)\}_{j=1}^\infty$ is a Cauchy sequence in Y . Moreover, since Y is a Banach space, $\{\mathsf{T}_j(x)\}_{j=1}^\infty$ converges to $y \in Y$. Accordingly, $y = \lim_{j \rightarrow \infty} \mathsf{T}_j(x) = \mathsf{T}(x)$. Clearly, T is linear, and $\|\mathsf{T}(x)\|_Y \leq C\|x\|_X$ for all $x \in X$. This means $\mathsf{T} \in \mathcal{B}(X, Y)$.

Note that we have not yet proved that $\{\mathsf{T}_j\}_{j=1}^\infty$ converges to T in norm. But, since $\{\mathsf{T}_j\}_{j=1}^\infty$ is a Cauchy sequence, for every $\varepsilon > 0$ there is an $n_0 \in \mathbf{N}$ such that

$$j > k > n_0 \Rightarrow \|\mathsf{T}_j - \mathsf{T}_k\| < \varepsilon.$$

Consequently,

$$j > k > n_0 \Rightarrow \|\mathsf{T}_j(x) - \mathsf{T}_k(x)\|_Y \leq \varepsilon \|x\|_X \text{ for all } x \in X.$$

If $j \rightarrow \infty$, then

$$k > n_0 \Rightarrow \|\mathsf{T}(x) - \mathsf{T}_k(x)\|_Y \leq \varepsilon \|x\|_X.$$

That is to say, $\|\mathsf{T}_k - \mathsf{T}\| \leq \varepsilon$ as $k > n_0$. Thus, $\|\mathsf{T}_k - \mathsf{T}\| \rightarrow 0$ as $k \rightarrow \infty$. \square

Theorem 9.3.1 particularly says that the dual space $X^* = \mathcal{B}(X, \mathbf{F})$ of the normed linear space X over a given field \mathbf{F} is always a Banach space no matter whether X is complete or not. To see this picture clearly, we will identify the dual spaces of three Banach spaces. Thanks to their importance, we will state them as three independent theorems.

The first one is about the sequence spaces.

Theorem 9.3.2. *Let $p \in [1, \infty)$. Then $\ell_p(\mathbf{N}, \mathbf{F})^* = \ell_{p(p-1)^{-1}}(\mathbf{N}, \mathbf{F})$ in the sense that $L \in \ell_p(\mathbf{N}, \mathbf{F})^*$ if and only if there is a unique $y = (y_k)_{k=1}^\infty \in \ell_{p(p-1)^{-1}}(\mathbf{N}, \mathbf{F})$ such that*

$$L(x) = \sum_{k=1}^{\infty} x_k \overline{y_k} \text{ for all } x = (x_k)_{k=1}^\infty \in \ell_p(\mathbf{N}, \mathbf{F}).$$

Proof. Since

$$\sum_{k=1}^{\infty} |x_k y_k| \leq \|x\|_p \|y\|_{p(p-1)^{-1}},$$

we have that the above-defined functional L is an element in $\ell_p(\mathbf{N}, \mathbf{F})^*$. Conversely, assume $L \in \ell_p(\mathbf{N}, \mathbf{F})^*$. To get the representation of L , we consider two cases.

Case 1: $p \in (1, \infty)$. Suppose that $x = (x_k)_{k=1}^\infty \in \ell_p(\mathbf{N}, \mathbf{F})$. For each $j \in \mathbf{N}$ let e_j be the vector in $\ell_p(\mathbf{N}, \mathbf{F})$ having the j th entry equal to 1 and all other entries equal to 0. Set $z_n = \sum_{j=1}^n x_j e_j$. Then $z_n \in \ell_p(\mathbf{N}, \mathbf{F})$ and

$$\lim_{n \rightarrow \infty} \|x - z_n\|_p = \lim_{n \rightarrow \infty} \left(\sum_{j=n+1}^\infty |x_j|^p \right)^{\frac{1}{p}} = 0.$$

Consequently,

$$L(z_n) = \sum_{j=1}^n x_j L(e_j)$$

and

$$|L(x) - L(z_n)| = |L(x - z_n)| \leq \|L\| \|x - z_n\|_p \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence

$$L(x) = \sum_{j=1}^\infty x_j L(e_j).$$

Let $y_j = \overline{L(e_j)}$ and $y = (y_j)_{j=1}^\infty$. Then

$$L(x) = \sum_{j=1}^\infty x_j \overline{y_j},$$

and hence it remains to show $y \in \ell_{p(p-1)^{-1}}(\mathbf{N}, \mathbf{F})$. To this end, we choose

$$x_j = \begin{cases} |y_j|^{p(p-1)^{-1}-2} y_j, & y_j \neq 0 \\ 0, & y_j = 0 \end{cases}.$$

This choice gives

$$\|z_n\|_p^p = \sum_{j=1}^n |x_j|^p = \sum_{j=1}^n |y_j|^{p(p-1)^{-1}}.$$

Moreover

$$L(z_n) = \sum_{j=1}^n |y_j|^{p(p-1)^{-1}} \text{ and } |L(z_n)| \leq \|L\| \|z_n\|_p = \|L\| \left(\sum_{j=1}^n |y_j|^{p(p-1)^{-1}} \right)^{\frac{1}{p}}.$$

Hence

$$\left(\sum_{j=1}^n |y_j|^{p(p-1)^{-1}} \right)^{(p-1)p^{-1}} \leq \|L\|.$$

This yields $\|y\|_{p(p-1)^{-1}} \leq \|L\| < \infty$, and so that for any $x \in \ell_p(\mathbf{N}, \mathbf{F})$,

$$|L(x)| = \left| \sum_{j=1}^{\infty} x_j \overline{y_j} \right| \leq \|x\|_p \|y\|_{p(p-1)^{-1}}.$$

Accordingly, $\|L\| = \|y\|_{p(p-1)^{-1}}$. The uniqueness of y is obvious.

Case 2: $p = 1$. To handle this case, we follow the foregoing argument until the definition of the vector y . To prove $y \in \ell_{\infty}(\mathbf{N}, \mathbf{F})$, we just observe that for each $j \in \mathbf{N}$,

$$|y_j| = |L(e_j)| \leq \|L\| \|e_j\|_1 = \|L\|.$$

Additionally, we obtain that for any $x \in \ell_1(\mathbf{N}, \mathbf{F})$,

$$|L(x)| = \left| \sum_{j=1}^{\infty} x_j \overline{y_j} \right| \leq \|y\|_{\infty} \|x\|_1.$$

Thus, $\|L\| = \|y\|_{\infty}$. Of course, y is uniquely determined. □

Remark 9.3.3.

(i) *From the argument for Theorem 9.3.2 it follows that $(\mathbf{F}^n)^* = \mathbf{F}^n$, i.e., $L \in (\mathbf{F}^n)^*$ if and only if there exists a unique $y = (y_1, \dots, y_n) \in \mathbf{F}^n$ such that*

$$L(x) = \sum_{k=1}^n x_k \overline{y_k} \text{ for all } x = (x_1, \dots, x_n) \in \mathbf{F}^n.$$

(ii) *If $p \in (1, \infty)$, then $\ell_p(\mathbf{N}, \mathbf{F})$ is reflexive. However, $\ell_1(\mathbf{N}, \mathbf{F})$ is not reflexive. To see this, since $\ell_1(\mathbf{N}, \mathbf{F})^* = \ell_{\infty}(\mathbf{N}, \mathbf{F})$, it is enough to prove $\ell_{\infty}(\mathbf{N}, \mathbf{F})^* \neq \ell_1(\mathbf{N}, \mathbf{F})$.*

As a matter of fact, if $c_0(\mathbf{N}, \mathbf{F})$ stands for the space of all vectors $(x_j)_{j=1}^{\infty}$ in $\ell_{\infty}(\mathbf{N}, \mathbf{F})$ with $\lim_{j \rightarrow \infty} x_j = 0$ then

$$c_0(\mathbf{N}, \mathbf{F}) \neq \ell_{\infty}(\mathbf{N}, \mathbf{F}) \text{ and } c_0(\mathbf{N}, \mathbf{F})^* = \ell_1(\mathbf{N}, \mathbf{F})$$

whose equality can be verified in a similar way to proving Theorem 9.3.2 with $p = 1$, and hence $c_0(\mathbf{N}, \mathbf{F})$ is not reflexive. Moreover, there exists a bounded linear operator $L : \ell_1(\mathbf{N}, \mathbf{F}) \rightarrow c_0(\mathbf{N}, \mathbf{F})^$ determined by*

$$L(x, y) = \sum_{j=1}^{\infty} x_j \overline{y_j} \text{ for } x = (x_j)_{j=1}^{\infty} \in \ell_1(\mathbf{N}, \mathbf{F}) \text{ and } y = (y_j)_{j=1}^{\infty} \in c_0(\mathbf{N}, \mathbf{F}).$$

Consequently, L maps $\ell_1(\mathbf{N}, \mathbf{F})$ to $\ell_\infty(\mathbf{N}, \mathbf{F})^*$ with $\|L(\cdot, y)\|_{\ell_\infty(\mathbf{N}, \mathbf{F})^*} \leq \|y\|_1$. Note that $c_0(\mathbf{N}, \mathbf{F})$ is a closed subspace of $\ell_\infty(\mathbf{N}, \mathbf{F})$. So

$$\|y\|_1 \leq \|L(\cdot, y)\|_{\ell_\infty(\mathbf{N}, \mathbf{F})^*} \leq \|y\|_1.$$

According to Corollary 8.4.4(ii), there is $0 \neq F \in \ell_\infty(\mathbf{N}, \mathbf{F})^*$ such that $F(y) = 0$ for all $y \in c_0(\mathbf{N}, \mathbf{F})$. This, along with the last estimation, implies $F \notin \ell_1(\mathbf{N}, \mathbf{F})^*$ - namely - there is no element $(y_j)_{j=1}^\infty \in \ell_1(\mathbf{N}, \mathbf{F})^*$ satisfying

$$F(x) = \sum_{j=1}^\infty x_j \overline{y_j} \text{ for all } x = (x_j)_{j=1}^\infty \in \ell_\infty(\mathbf{N}, \mathbf{F}).$$

The second one characterizes the Banach dual of the space of all \mathbf{F} -valued continuous functions on a given finite closed interval in \mathbf{R} .

Theorem 9.3.4. For $-\infty < a < b < \infty$ equip $f \in C([a, b], \mathbf{F})$ with the sup-norm

$$\|f\|_\infty = \max_{x \in [a, b]} |f(x)|,$$

and let $BV([a, b], \mathbf{F})$ be the class of all \mathbf{F} -valued functions having bounded variation on $[a, b]$. If

$$BV_0([a, b], \mathbf{F}) = \left\{ g \in BV([a, b], \mathbf{F}) : \int_a^b f(x) d\overline{g(x)} = 0 \text{ for all } f \in C([a, b], \mathbf{F}) \right\},$$

then

$$C([a, b], \mathbf{F})^* = BV([a, b], \mathbf{F}) / BV_0([a, b], \mathbf{F})$$

in the sense that $L \in C([a, b], \mathbf{F})^*$ when and only when there is a unique equivalent class

$$g + BV_0([a, b], \mathbf{F}) \in BV([a, b], \mathbf{F}) / BV_0([a, b], \mathbf{F}),$$

such that L can be written as the Riemann-Stieltjes integral

$$L(f) = \int_a^b f(x) d\overline{g(x)} \text{ for all } f \in C([a, b], \mathbf{F}).$$

Proof. On the one hand, suppose that the statement after the when and only when is true. We may write $g = g_1 + ig_2$ where $g_1, g_2 \in BV[a, b]$, i.e., $V_a^b(g_k) < \infty$

for $k = 1, 2$, in the sense of Definition 3.1.1, and consequently,

$$\begin{aligned} |L(f)| &= \left| \int_a^b f(x) d(g_1(x) - ig_2(x)) \right| \\ &= \left| \int_a^b f(x) dg_1(x) - i \int_a^b f(x) dg_2(x) \right| \\ &\leq \|f\|_\infty (V_a^b(g_1) + V_a^b(g_2)). \end{aligned}$$

Thus, this functional L belongs to $C([a, b], \mathbf{F})^*$.

Conversely, suppose that $L \in C([a, b], \mathbf{F})^*$ and \mathbf{E} is a Hahn-Banach extension of L to the space $BD([a, b], \mathbf{F})$ of all functions $f : [a, b] \rightarrow \mathbf{F}$ satisfying $\|f\|_\infty = \sup\{|f(x)| : x \in [a, b]\} < \infty$. For $t \in [a, b]$ let $g(t) = \overline{\mathbf{E}(1_{[a, t]})}$, but also for $\{t_0, t_1, \dots, t_n\}$, a partition of $[a, b]$, and for each $k \in \{1, \dots, n\}$ let λ_k be the sign of $g(t_k) - g(t_{k-1})$, i.e., the number in \mathbf{F} obeying

$$|g(t_k) - g(t_{k-1})| = \lambda_k (g(t_k) - g(t_{k-1})).$$

Because of the following representation of 1:

$$1 = \sum_{k=1}^n (1_{[a, t_k]} - 1_{[a, t_{k-1}]}) \quad \text{on } [a, b],$$

one has

$$\begin{aligned} \sum_{k=1}^n |g(t_k) - g(t_{k-1})| &= \sum_{k=1}^n \overline{\lambda_k (\mathbf{E}(1_{[a, t_k]}) - \mathbf{E}(1_{[a, t_{k-1}]}))} \\ &\leq \left| \mathbf{E} \left(\sum_{k=1}^n \overline{\lambda_k} (1_{[a, t_k]} - 1_{[a, t_{k-1}]}) \right) \right| \\ &\leq \|\mathbf{E}\| \left\| \sum_{k=1}^n \overline{\lambda_k} (1_{[a, t_k]} - 1_{[a, t_{k-1}]}) \right\|_\infty \\ &\leq \|\mathbf{E}\| \left\| \sum_{k=1}^n (1_{[a, t_k]} - 1_{[a, t_{k-1}]}) \right\|_\infty \\ &\leq \|\mathbf{E}\|. \end{aligned}$$

This means that g has bounded variation on $[a, b]$. Now let $f \in C([a, b], \mathbf{F})$ and $t_k = a + kn^{-1}(b-a)$ for $k = 0, \dots, n$, and define $f_n \in BD([a, b], \mathbf{F})$ via

$$f_n(x) = \sum_{k=1}^n f(t_{k-1}) (1_{[a, t_k]}(x) - 1_{[a, t_{k-1}]}(x)) \quad \text{for all } x \in [a, b].$$

Now, it follows from the foregoing representation of 1 and the uniform continuity of f on $[a, b]$ that

$$\|f_n - f\|_\infty \leq \max_{1 \leq k \leq n} \sup_{x \in [t_{k-1}, t_k]} |f(x) - f(t_{k-1})| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Note that E is continuous on $BD([a, b], \mathbf{F})$. So it follows from the definition of the Riemann-Stieltjes integral that

$$\begin{aligned} L(f) &= E(f) \\ &= \lim_{n \rightarrow \infty} E(f_n) \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n f(t_{k-1}) \overline{(g(t_k) - g(t_{k-1}))} \\ &= \int_a^b f(x) d\overline{g(x)}, \end{aligned}$$

as required. Finally, if

$$h + BV_0([a, b], \mathbf{F}) \in BV([a, b], \mathbf{F}) / BV_0([a, b], \mathbf{F})$$

also satisfies

$$L(f) = \int_a^b f(x) d\overline{h(x)} \text{ for all } f \in C([a, b], \mathbf{F}),$$

then

$$\int_a^b f(x) d\overline{(g(x) - h(x))} = 0 \text{ for all } f \in C([a, b], \mathbf{F}),$$

and hence $g - h \in BV_0([a, b], \mathbf{F})$, as desired. \square

To reach the third dual Banach space, we need the following Olof Hanner's inequality which is viewed as a sort of improvement of the Minkowski inequality.

Theorem 9.3.5. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing, $E \subseteq \mathbf{R}$ m_g -measurable, $p \in [1, \infty)$, and $f_1, f_2 \in LRS_g^p(E, \mathbf{F})$.*

(i) *If $p \in [1, 2]$, then*

$$\begin{aligned} &\|f_1 + f_2\|_{p, m_g, E}^p + \|f_1 - f_2\|_{p, m_g, E}^p \\ &\geq (\|f_1\|_{p, m_g, E} + \|f_2\|_{p, m_g, E})^p + \left| \|f_1\|_{p, m_g, E} - \|f_2\|_{p, m_g, E} \right|^p \end{aligned}$$

and

$$\begin{aligned} & (\|f_1 + f_2\|_{p,m_g,E} + \|f_1 - f_2\|_{p,m_g,E})^p + \left| \|f_1 + f_2\|_{p,m_g,E} - \|f_1 - f_2\|_{p,m_g,E} \right|^p \\ & \leq 2^p (\|f_1\|_{p,m_g,E}^p + \|f_2\|_{p,m_g,E}^p). \end{aligned}$$

(ii) If $p \in [2, \infty)$, then

$$\begin{aligned} & \|f_1 + f_2\|_{p,m_g,E}^p + \|f_1 - f_2\|_{p,m_g,E}^p \\ & \leq (\|f_1\|_{p,m_g,E} + \|f_2\|_{p,m_g,E})^p + \left| \|f_1\|_{p,m_g,E} - \|f_2\|_{p,m_g,E} \right|^p \end{aligned}$$

and

$$\begin{aligned} & (\|f_1 + f_2\|_{p,m_g,E} + \|f_1 - f_2\|_{p,m_g,E})^p + \left| \|f_1 + f_2\|_{p,m_g,E} - \|f_1 - f_2\|_{p,m_g,E} \right|^p \\ & \geq 2^p (\|f_1\|_{p,m_g,E}^p + \|f_2\|_{p,m_g,E}^p). \end{aligned}$$

Proof. It suffices to verify the first inequalities in (i) and (ii) since the second ones follow from the first ones via replacing f_1 and f_2 with $f_1 + f_2$ and $f_1 - f_2$. Note that the cases $p = 1$ and $p = 2$ are clear from a direct computation. So, it remains to check the case $p \neq 2$. To this end, we may assume

$$R = \frac{\|f_2\|_{p,m_g,E}}{\|f_1\|_{p,m_g,E}} \quad \text{and} \quad \|f_1\|_{p,m_g,E} = 1.$$

If $R \leq 1$, then we consider

$$[0, 1] \ni r \mapsto \begin{cases} \phi(r) = (1+r)^{p-1} + (1-r)^{p-1} \\ \psi(r) = r^{1-p}((1+r)^{p-1} - (1-r)^{p-1}) \end{cases},$$

and compute

$$\frac{d\phi(r)}{dr} + \left(\frac{d\psi(r)}{dr} \right) R^p = (p-1)((1+r)^{p-2} - (1-r)^{p-2}) \left(1 - \left(\frac{R}{r} \right)^p \right),$$

which vanishes only at $r = R$ and indicates that $\phi(r) + \psi(r)R^p$ has a maximum or minimum at $r = R$ when $p < 2$ or $p > 2$.

If $R > 1$, then

$$\psi(r) \begin{cases} \leq \phi(r), & p < 2 \\ \geq \phi(r), & p > 2 \end{cases} \Rightarrow \phi(r) + \psi(r)R^p \begin{cases} \leq \psi(r) + \phi(r)R^p, & p < 2 \\ \geq \psi(r) + \phi(r)R^p, & p > 2 \end{cases}.$$

Accordingly, for $|a|, |b| \geq 0$ one gets

$$\phi(r)|a|^p + \psi(r)|b|^p \begin{cases} \leq |a+b|^p + |a-b|^p, & p < 2 \\ \geq |a+b|^p + |a-b|^p, & p > 2 \end{cases},$$

with equality when $r = a^{-1}b \leq 1$, $a > 0$, $b \geq 0$. This in turn implies

$$\int_E (\phi(r)|f_1|^p dm_g + \psi(r)|f_2|^p) dm_g \begin{cases} \leq \int_E (|f_1 + f_2|^p + |f_1 - f_2|^p) dm_g, & p < 2 \\ \geq \int_E (|f_1 + f_2|^p + |f_1 - f_2|^p) dm_g, & p > 2 \end{cases}.$$

Of course, the last estimates, along with $r = R$, give the desired inequalities right away. \square

An immediate example of the use of Theorem 9.3.5 is given by the following projection result which will be used later on.

Corollary 9.3.6. *Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be increasing, $E \subseteq \mathbf{R}$ m_g -measurable, $p \in (1, \infty)$, and \mathcal{K} a closed convex subset of $LRS_g^p(E)$. If $f \in LRS_g^p(E, \mathbf{F}) \setminus \mathcal{K}$, then there exists a function $h_0 \in \mathcal{K}$ such that*

$$\|f - h_0\|_{p, m_g, E} = \inf_{h \in \mathcal{K}} \|f - h\|_{p, m_g, E}.$$

Moreover, every function $h \in \mathcal{K}$ satisfies

$$\Re \int_E (h - h_0) \overline{(f - h_0)} |f - h_0|^{p-2} dm_g \leq 0.$$

Proof. Without loss of generality, we may assume and $f = 0$ and $\delta = \inf_{h \in \mathcal{K}} \|h\|_{p, m_g, E}$.

For the first result, suppose that $\{h_j\}_{j=1}^\infty$ is in \mathcal{K} with $\lim_{j \rightarrow \infty} \|h_j\|_{p, m_g, E} = \delta$. Since \mathcal{K} is convex, $2^{-1}(h_j + h_k) \in \mathcal{K}$ and consequently,

$$\begin{aligned} 2\delta &\leq \|h_j + h_k\|_{p, m_g, E} \\ &\leq \|h_j\|_{p, m_g, E} + \|h_k\|_{p, m_g, E} \\ &\rightarrow 2\delta \text{ as } k, j \rightarrow \infty. \end{aligned}$$

Suppose that $\{h_n\}_{n=1}^\infty$ is not a Cauchy sequence in $LRS_g^p(E, \mathbf{F})$. Then there is a constant $\kappa > 0$ such that $\|h_j - h_k\|_{p, m_g, E} \geq \kappa$ holds for infinitely many j 's and k 's.

Case 1: $p \in (1, 2]$. According to the second inequality in Theorem 9.3.5 (i), we have

$$\begin{aligned} & (\|h_j + h_k\|_{p, m_g, E} + \|h_j - h_k\|_{p, m_g, E})^p + \left| \|h_j + h_k\|_{p, m_g, E} - \|h_j - h_k\|_{p, m_g, E} \right|^p \\ & \leq 2^p (\|h_j\|_{p, m_g, E}^p + \|h_k\|_{p, m_g, E}^p) \\ & \rightarrow 2^{p+1} \delta^p \text{ as } j, k \rightarrow \infty, \end{aligned}$$

whence getting

$$|2\delta + \kappa|^p + |2\delta - \kappa|^p \leq 2^{p+1} \delta^p.$$

Since $\phi(x) = |2\delta + x|^p$ is a strictly convex function with $x \in \mathbf{R}$; that is to say,

$$\phi(tx_1 + (1-t)x_2) < t\phi(x_1) + (1-t)\phi(x_2) \text{ for all } (t, x_1, x_2) \in (0, 1) \times \mathbf{R} \times \mathbf{R},$$

we obtain the following contradiction:

$$2^{1+p} \delta^p < |2\delta + \kappa|^p + |2\delta - \kappa|^p \leq 2^{p+1} \delta^p.$$

Case 2: $p \in [2, \infty)$. According to the first inequality in Theorem 9.3.5 (ii), we have

$$\begin{aligned} & \|h_j + h_k\|_{p, m_g, E}^p + \|h_j - h_k\|_{p, m_g, E}^p \\ & \leq (\|h_j\|_{p, m_g, E} + \|h_k\|_{p, m_g, E})^p + \left| \|h_j\|_{p, m_g, E} - \|h_k\|_{p, m_g, E} \right|^p, \end{aligned}$$

whence deriving a contradiction

$$(2\delta)^p < (2\delta)^p + \kappa^p \leq (2\delta)^p.$$

Therefore, $\{h_n\}_{n=1}^\infty$ must be a Cauchy sequence in $LRS_g^p(E, \mathbf{F})$. Note that \mathcal{K} is closed. So $\{h_n\}_{n=1}^\infty$ is convergent to an element $h_0 \in \mathcal{K}$ with $\delta = \|h_0\|_{p, m_g, E}$.

For the second result, let $h \in \mathcal{K}$. Since \mathcal{K} is convex, we have

$$h_t = th + (1-t)h_0 \in \mathcal{K} \text{ for all } t \in [0, 1],$$

and consequently,

$$\phi(t) = \|h_t\|_{p, m_g, E}^p \geq \delta^p \text{ while } \phi(0) = \delta^p.$$

Noticing both

$$\lim_{t \rightarrow 0^+} t^{-1} (|h_t|^p - |h_0|^p) = 2^{-1} p |h_0|^{p-2} ((\overline{h-h_0})h_0 + (h-h_0)\overline{h_0})$$

and

$$|h_0|^p - |2h_0 - h|^p \leq t^{-1} (|h_t|^p - |h_0|^p) \leq |h|^p - |h_0|^p$$

which follows from the convexity

$$(tx_1 + (1-t)x_2)^p \leq tx_1^p + (1-t)x_2^p$$

of the function x^p on $[0, \infty)$, we use the dominated convergence theorem to obtain that $\phi'(0)$ exists with

$$0 \leq \phi'(0) = \Re \int_E (h - h_0) \overline{h_0} |h_0|^{p-2} dm_g,$$

as desired. \square

Now, we are ready to identify the dual space of each $LRS_g^p(E, \mathbf{F})$ for $p \in [1, \infty)$ via the Frigyes Riesz representation theorem below.

Theorem 9.3.7. *Let $g: \mathbf{R} \rightarrow \mathbf{R}$ be increasing, $E \subseteq \mathbf{R}$ m_g -measurable and $p \in [1, \infty)$. Then*

$$LRS_g^p(E, \mathbf{F})^* = LRS_g^{p(p-1)^{-1}}(E, \mathbf{F})$$

in the sense that $L \in LRS_g^p(E, \mathbf{F})^$ if and only if there is a unique element $h \in LRS_g^{p(p-1)^{-1}}(E, \mathbf{F})$ such that*

$$L(f) = \int_E f \overline{h} dm_g \text{ for all } f \in LRS_g^p(E, \mathbf{F}).$$

Proof. The Hölder inequality implies the above-formulated functional is an element of $LRS_g^p(E, \mathbf{F})^*$.

For the reversed conclusion, we may assume $0 \neq L \in LRS_g^p(E, \mathbf{F})^*$. Then we consider two cases as follows.

Case 1: $p \in (1, \infty)$. For the above-given L , define

$$\mathcal{K} = \{f \in LRS_g^p(E, \mathbf{F}) : L(f) = 0\}.$$

Since L is continuous, the subset \mathcal{K} of $LRS_g^p(E, \mathbf{F})$ is closed and convex. Because of $L \neq 0$, there exists an element $f_0 \in LRS_g^p(E, \mathbf{F})$ such that $L(f_0) \neq 0$, namely, $f_0 \notin \mathcal{K}$. According to Corollary 9.3.6 and the linearity of \mathcal{K} , we can find an $h_0 \in \mathcal{K}$ such that

$$\Re \int_E (h - h_0) |f_0 - h_0|^{p-2} (\overline{f_0 - h_0}) dm_g \leq 0 \text{ for all } h \in \mathcal{K}.$$

Choosing $h = 2h_0$ in the above estimate, we get

$$\Re \int_E h_0 |f_0 - h_0|^{p-2} (\overline{f_0 - h_0}) dm_g \leq 0$$

whence finding

$$\Re \int_E (\pm h) |f_0 - h_0|^{p-2} (\overline{f_0 - h_0}) dm_g \leq 0 \text{ for all } h \in \mathcal{K}.$$

Note that if $\mathbf{F} = \mathbf{C}$ then $ih \in \mathcal{K}$ and hence the last inequality still holds when $\pm h$ is replaced by ih . So, the last integral equals 0 for all $h \in \mathcal{K}$.

If f is an arbitrary element of $LRS_g^p(E, \mathbf{F})$ with

$$f = f_1 + f_2 \text{ and } f_1 = \left(\frac{L(f)}{L(f_0 - h_0)} \right) (f_0 - h_0),$$

then $f_2 \in \mathcal{K}$ and hence

$$\begin{aligned} & \int_E f |f_0 - h_0|^{p-2} (\overline{f_0 - h_0}) dm_g \\ &= \int_E f_1 |f_0 - h_0|^{p-2} (\overline{f_0 - h_0}) dm_g + \int_E f_2 |f_0 - h_0|^{p-2} (\overline{f_0 - h_0}) dm_g \\ &= \int_E f_1 |f_0 - h_0|^{p-2} (\overline{f_0 - h_0}) dm_g \\ &= L(f) \left(\frac{\int_E |f_0 - h_0|^p dm_g}{L(f_0 - h_0)} \right). \end{aligned}$$

Clearly, the desired element h determining the representation of L is

$$h = \frac{|f_0 - h_0|^{p-2} (f_0 - h_0) \overline{L(f_0 - h_0)}}{\int_E |f_0 - h_0|^p dm_g}.$$

In fact, such an element h is unique – if there is another element \tilde{h} then

$$\int_E \overline{(h - \tilde{h})} f dm_g = 0 \text{ for } f = (h - \tilde{h}) |h - \tilde{h}|^{\frac{2-p}{p-1}} \in LRS_g^p(E, \mathbf{F})$$

and hence $h = \tilde{h}$ m_g -a.e. on E .

Case 2: $p = 1$. We handle this case via two situations. The first one is $m_g(E) < \infty$. Under this condition, L is also in $LRS_g^q(E, \mathbf{F})^*$ (for any $q \in (1, \infty)$) since the Hölder inequality yields that

$$|L(f)| \leq \|L\| \|f\|_{1, m_g, E} \leq \|L\| (m_g(E))^{\frac{q-1}{q}} \|f\|_{q, m_g, E}$$

is valid for all $f \in LRS_g^q(E, \mathbf{F})$. However, according to the previous *Case 1* there exists a unique $h \in LRS_g^{\frac{q}{q-1}}(E, \mathbf{F})$ such that

$$L(f) = \int_E f \bar{h} dm_g \quad \text{for all } f \in LRS_g^q(E, \mathbf{F}).$$

So, by the inclusion $LRS_g^{p_2}(E, \mathbf{F}) \subseteq LRS_g^{p_1}(E, \mathbf{F})$, which holds for $1 < p_1 < p_2 < \infty$ and follows from Hölder's inequality, indicates that h does not depend on q . Now if $f = |h|^{\frac{q}{q-1}-2} \bar{h}$, then $f \in LRS_g^q(E, \mathbf{F})$ and hence

$$\|h\|_{\frac{q}{q-1}, m_g, E}^{\frac{q}{q-1}} = L(f) \leq \|L\| (m_g(E))^{\frac{q-1}{q}} \|h\|_{\frac{q}{q-1}, m_g, E}^{\frac{q}{q-1}-1}.$$

Accordingly, for any $\varepsilon > 0$ we have

$$(\|L\| + \varepsilon) \left(m_g(\{x \in E : |h(x)| > \|L\| + \varepsilon\}) \right)^{\frac{q-1}{q}} \leq \|h\|_{\frac{q}{q-1}, m_g, E}^{\frac{q}{q-1}} \leq \|L\| (m_g(E))^{\frac{q-1}{q}}.$$

Since p can be sufficiently large, the last estimate forces

$$m_g(\{x \in E : |h(x)| > \|L\| + \varepsilon\}) = 0,$$

namely, $h \in LRS_g^\infty(E, \mathbf{F})$. This fact further implies that $\int_E |h| |f| dm_g < \infty$ for all $f \in LRS_g^1(E, \mathbf{F})$. With $f \in LRS_g^1(E, \mathbf{F})$ and $j \in \mathbf{N}$, we consider

$$f_j(x) = \begin{cases} f(x) & , \quad |f(x)| \leq j \\ 0 & , \quad |f(x)| > j \end{cases},$$

and find that

$$\begin{cases} |f_j| \in LRS_g^q(E); \\ f_j(x) \rightarrow f(x); \\ |f_j(x)| \leq |f(x)|. \end{cases}$$

By the dominated convergence theorem it follows that

$$\begin{cases} \lim_{j \rightarrow \infty} \|f_j - f\|_{1, m_g, E} = 0; \\ \lim_{j \rightarrow \infty} \|hf_j - hf\|_{1, m_g, E} = 0; \\ L(f) = \lim_{j \rightarrow \infty} L(f_j) = \lim_{j \rightarrow \infty} \int_E f_j \bar{h} dm_g = \int_E f \bar{h} dm_g. \end{cases}$$

The second one is $m_g(E) = \infty$. Since

$$E = \bigcup_{j=-\infty}^{\infty} E \cap [j, j+1) \quad \text{and} \quad m_g([j, j+1)) < \infty,$$

any function $f \in LRS_g^1(E, \mathbf{F})$ can be written as

$$f(x) = \sum_{j=-\infty}^{\infty} f(x) 1_{E \cap [j, j+1)}(x) \text{ for all } x \in E.$$

Note that $f \in LRS_g^1(E \cap [j, j+1), \mathbf{F})$ if and only if $f 1_{E \cap [j, j+1)} \in LRS_g^1(E, \mathbf{F})$. So, $L_j(f) = L(f 1_{E \cap [j, j+1)})$ defines an element in $LRS_g^1(E \cap [j, j+1), \mathbf{F})^*$, and consequently, there exists an element $h_j \in LRS_g^\infty(E \cap [j, j+1), \mathbf{F})$ such that

$$L(f 1_{E \cap [j, j+1)}) = \int_{E \cap [j, j+1)} f \bar{h}_j dm_g \text{ and } \|h_j\|_{\infty, m_g, E \cap [j, j+1)} \leq \|L_j\| \leq \|L\|.$$

Now, if the function h is defined on E by setting $h = h_j$ on $E \cap [j, j+1)$ for each $j \in \mathbf{Z}$, then $h \in LRS_g^\infty(E, \mathbf{F})$ and

$$L(f) = L\left(\sum_{j=-\infty}^{\infty} f 1_{E \cap [j, j+1)}\right) = \sum_{j=-\infty}^{\infty} \int_{E \cap [j, j+1)} f \bar{h}_j dm_g = \int_E f \bar{h} dm_g,$$

owing to the countable additivity of the measure m_g . Moreover, if there is another function \tilde{h} satisfying the last representation, then $\int_E |h - \tilde{h}| |f| dm_g = 0$ for all $f \in LRS_g^1(E, \mathbf{F})$, and hence for $f = 1_{E \cap [j, j+1)}$ where $j \in \mathbf{Z}$. This deduces $h = \tilde{h}$ m_g -a.e. on E . Therefore, the function h is uniquely determined. \square

Remark 9.3.8. On the one hand, $LRS_g^{1 < p < \infty}(E, \mathbf{F})$ is reflexive, but $LRS_g^1(E, \mathbf{F})$ is not reflexive generally - there exists a linear continuous functional $0 \neq T : LRS_g^\infty(E, \mathbf{R}) \rightarrow \mathbf{R}$ (generated by an $h \in LRS_g^\infty(E, \mathbf{F})$) which cannot be written as

$$T(f) = \int_E f \bar{h} dm_g \text{ for all } f \in LRS_g^1(E, \mathbf{F}).$$

To see the last assertion, let us take $g(x) = x$, $E = \mathbf{R} = \mathbf{F}$ and W be the subspace of all continuous functions in $LRS_{id}^\infty(\mathbf{R}, \mathbf{R})$. Upon defining the linear operator $T : W \ni f \mapsto T(f) = f(0) \in \mathbf{R}$, we see that $|T(f)| = |f(0)| \leq \|f\|_{\infty, m_{id}, \mathbf{R}}$ and so that T is continuous with respect to $\|\cdot\|_{\infty, m_{id}, \mathbf{R}}$. The Hahn-Banach extension theorem ensures that T can be extended to $\bar{T} \in (LRS_{id}^\infty(\mathbf{R}, \mathbf{R}))^*$. Assume now that

$$\bar{T}(f) = \int_{\mathbf{R}} f h dm_{id} \text{ for some } h \in LRS_{id}^1(\mathbf{R}, \mathbf{R}) \text{ and for all } f \in LRS_{id}^\infty(\mathbf{R}, \mathbf{R}).$$

Then

$$\mathsf{T}(f) = \int_{\mathbf{R}} f h d m_{id} = f(0) \text{ for all } f \in W.$$

Observe that for $k \in \mathbf{N}$ and $x \neq 0$ with $\varepsilon < |x|$ there is continuous function f_k with support in $[x - \varepsilon, x + \varepsilon]$ which converges in $LRS_{id}^\infty(\mathbf{R}, \mathbf{R})$ to $1_{[x - \varepsilon, x + \varepsilon]}$. Thus, we find

$$0 = (2\varepsilon)^{-1} \lim_{k \rightarrow \infty} \int_{\mathbf{R}} h f_k d m_{id} = (2\varepsilon)^{-1} \int_{[x - \varepsilon, x + \varepsilon]} h d m_{id}.$$

However, if x is the Lebesgue point of h , then letting $\varepsilon \rightarrow 0$ derives $h(x) = 0$, and hence $h = 0$ almost everywhere on \mathbf{R} - and yet - we should conclude $\mathsf{T} = 0$ which is a contradiction.

On the other hand, from Corollary 9.3.6 to Theorem 9.3.7 we have used the following definition of integration

$$\int_E f d m_g = \int_E (\Re f + i \Im f) d m_g = \int_E \Re f d m_g + i \int_E \Im f d m_g$$

(for any $|f| \in LRS_g^1(E, \mathbf{F})$) and its linearity

$$\int_E (c_1 f_1 + c_2 f_2) d m_g = c_1 \int_E f_1 d m_g + c_2 \int_E f_2 d m_g$$

(for any $|f_1|, |f_2| \in LRS_g^1(E, \mathbf{F})$ and $c_1, c_2 \in \mathbf{F}$) as well as some elementary algebraic operations of complex numbers.

9.4 Weak and Weak* Topologies

This section is about to discuss some special topologies on a given Banach space X and on its dual X^* , which are weaker than the usual ones.

Definition 9.4.1. Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} . Then a sequence $\{x_k\}_{k=1}^\infty$ in X is said to converge weakly to $x \in X$, denoted $x_k \xrightarrow{w} x$ provided that $\lim_{k \rightarrow \infty} |L(x_k) - L(x)| = 0$ for each $L \in X^*$.

Remark 9.4.2.

(i) Clearly, $x_k \rightarrow x$ in X implies $x_k \xrightarrow{w} x$, but not conversely - for example - consider $(\ell_2(\mathbf{N}, \mathbf{R}), \|\cdot\|_2)$. If $e_j = (\underbrace{0, \dots, 0}_{j-1}, 1, 0, \dots)$ for each $j \in \mathbf{N}$, then $\|e_j\|_2 = 1$

yields that $\{e_j\}_{j=1}^\infty$ does not converge strongly to 0; nevertheless according to Theorem 9.3.2, for any

$$L \in \mathcal{B}(\ell_2(\mathbf{N}, \mathbf{R}), \mathbf{R}) = \ell_2(\mathbf{N}, \mathbf{R})^*$$

there exists a $y = \{y_k\}_{k=1}^\infty \in \ell_2(\mathbf{N}, \mathbf{R})$ such that $\lim_{j \rightarrow \infty} L(e_j) = \lim_{j \rightarrow \infty} y_j = 0$, namely, $e_j \xrightarrow{w} 0$.

(ii) In addition, a sequence may have at most one weak limit for if $x \neq y$ then, as we saw in the proof of Corollary 8.4.4 (iv), there exists an element $L \in X^*$ such that

$$L(x) - L(y) = L(x - y) = 1.$$

If $x_k \xrightarrow{w} x$ and $x_k \xrightarrow{w} y$, then there is an $N \in \mathbf{N}$ such that $k > N$ implies

$$|L(x_k) - L(x)| < 2^{-1} \quad \text{and} \quad |L(x_k) - L(y)| < 2^{-1},$$

and hence $|L(x) - L(y)| < 1$, a contradiction.

The following is a natural way to introduce the weak topology where the basis is formed by all weakly open sets.

Definition 9.4.3. Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} .

(i) Given $\varepsilon > 0$ and a finite set $\{L_1, \dots, L_n\}$ in X^* , the $(\varepsilon, L_1, \dots, L_n)$ -neighborhood of a point $x \in X$, denoted $N(x; \varepsilon, L_1, \dots, L_n)$, is defined by

$$\{y \in X : |L_k(x) - L_k(y)| < \varepsilon, k = 1, \dots, n\}.$$

(ii) A subset of X is called weakly open provided that it is a union of the neighborhoods $N(x; \varepsilon, L_1, \dots, L_n)$. The complement of a weakly open set is called a weakly closed set.

(iii) A subset E of X is called weakly compact provided that every cover of E by weakly open sets has a finite subcover.

Remark 9.4.4. Note that $B_\varepsilon^X(x) \subseteq N(x; \varepsilon\|L\|, L)$ for each $L \in X^*$. So Definition 9.4.3 is a natural generalization of the idea of a ball in a metric space. From the definition it turns out that $x_k \xrightarrow{w} x$ is equivalent to saying that to any such weak neighborhood of x there corresponds an $n_0 \in \mathbf{N}$ such that x_k lies in this neighborhood for $k \geq n_0$.

Using the second dual we introduce the following concept.

Definition 9.4.5. Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} , and $L, L_1, L_2, \dots \in X^*$. Then we say that:

- (i) $\{L_n\}_{n=1}^\infty$ is strongly convergent to L provided $\lim_{n \rightarrow \infty} \|L_n - L\| = 0$;
- (ii) $\{L_n\}_{n=1}^\infty$ is weak* convergent to L provided $\lim_{n \rightarrow \infty} |L_n(x) - L(x)| = 0$ for all $x \in X$;
- (iii) $\{L_n\}_{n=1}^\infty$ is weakly convergent to L provided $\lim_{n \rightarrow \infty} \|F(L_n) - F(L)\| = 0$ for all $F \in X^{**}$.

Remark 9.4.6.

(i) In general, (ii) and (iii) in Definition 9.4.5 are not equivalent. But, if there is an isometry between X and X^{**} then both concepts coincide.

(ii) From Theorem 9.2.4 (ii) we can see that if $(X, \|\cdot\|_X)$ is a Banach space over \mathbf{F} , then any weak* convergent sequence in X^* must be bounded and its converse is also true whenever this sequence is assumed to be convergent on a dense subset of X .

(iii) Motivated by Definition 9.4.5 (iii), we say that $\{T_n\}_{n=1}^\infty$ in $\mathcal{B}(X, Y)$ is weakly convergent to $T \in \mathcal{B}(X, Y)$ provided

$$\lim_{n \rightarrow \infty} |F(T_n(x)) - F(T(x))| = 0 \text{ for all } (x, F) \in X \times Y^*.$$

Here $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are supposed to be normed linear spaces over \mathbf{F} . According to Definition 9.2.2, we can find out that the uniform convergence implies the strong convergence which implies the weak convergence, but not conversely. Since Example 9.2.3 has shown that the strong convergence does not yield the uniform convergence, it is enough to check that the weak convergence does not derive the strong convergence. To do so, consider $X = Y = \ell_2(\mathbf{N}, \mathbf{R})$ over \mathbf{R} , and

$$T_n(x) = (\underbrace{0, \dots, 0}_n, x_1, x_2, \dots) \text{ for } x = (x_1, x_2, \dots) \in \ell_2(\mathbf{N}, \mathbf{R}).$$

Clearly, $\|T_n\| = 1$. And, if $y = (y_1, y_2, \dots) \in \ell_2(\mathbf{N}, \mathbf{R})$, then

$$\left| \sum_{k=1}^{\infty} (T_n(x))_k y_k \right| = \left| \sum_{k=n+1}^{\infty} x_k y_k \right| \leq \|x\|_2 \left(\sum_{k=n+1}^{\infty} y_k^2 \right)^{2^{-1}} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Thus, from Example 9.3.2 it follows that $\{\mathsf{T}_n\}_{n=1}^\infty$ is weakly convergent to 0. But, $\{\mathsf{T}_n\}_{n=1}^\infty$ is not strongly convergent to 0 since

$$\|\mathsf{T}_n(e_1) - \mathsf{T}_m(e_1)\|_2 = \|e_{n+1} - e_{m+1}\|_2 = \sqrt{2}.$$

Similarly, we can define the weak-* topology on X^* via all weak* open subsets of X^* .

Definition 9.4.7. Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} .

(i) Given $\varepsilon > 0$ and a finite set $\{x_1, \dots, x_n\}$ in X , the $(\varepsilon, x_1, \dots, x_n)$ -neighborhood of a functional $F \in X^*$, denoted $N(F; \varepsilon, x_1, \dots, x_n)$, is defined by

$$\{G \in X^* : |G(x_k) - F(x_k)| < \varepsilon, k = 1, \dots, n\}.$$

(ii) An arbitrary union of such neighborhoods is called a weak* open set. The complement of a weak* open set is called a weak* closed set.

(iii) A subset A of X^* is called weak* compact provided that every cover of A by weak* open sets has a finite subcover.

Definition 9.4.7 yields immediately that a weak* closed subset of a weak* compact set is itself weak* compact. To grasp the essence of weak* compact sets, we extend Cartesian products from two sets to a family and work with their topological compactness.

Definition 9.4.8.

(i) A collection \mathcal{T} of subsets of a set X is said to be a topology for X provided that:

- (a) $\emptyset, X \in \mathcal{T}$;
- (b) $\bigcap_{j=1}^n E_j \in \mathcal{T}$ whenever $n \in \mathbf{N}$ and $E_1, E_2, \dots, E_n \in \mathcal{T}$;
- (c) $\bigcup_{j \in I} E_j \in \mathcal{T}$ whenever $\{E_j\}_{j \in I}$ is a family in \mathcal{T} .

In this case, (X, \mathcal{T}) is called a topological space. Meanwhile, members of \mathcal{T} and their complements are called open and closed sets of (X, \mathcal{T}) respectively. Furthermore, (X, \mathcal{T}) is called compact provided every cover of X by sets of \mathcal{T} has a finite subcover.

(ii) Let $\{(X_j, \mathcal{T}_j)\}_{j \in I}$ be a family of topological spaces indexed by the set I . Then the Cartesian product of $\{X_j\}_{j \in I}$ is defined by

$$X = \prod_{j \in I} X_j = \{(x_j)_{j \in I} : x_j \in X_j\}$$

Furthermore, for each $j \in I$ the projection $p_j : X \rightarrow X_j$ is defined by $p_j((x_\alpha)_{\alpha \in I}) = x_j$. The product topology \mathcal{T} on X is defined to be the topology with the fewest open sets for which every projection p_j is continuous; that is, $p_j^{-1}(O)$ is open in X whenever O is open in X_j . Consequently, the above-defined (X, \mathcal{T}) is called a product topological space.

Lemma 9.4.9. *Let \mathcal{F} be a family of subsets of a given set X . If \mathcal{F} enjoys the finite intersection property (FIP) – each finite collection of subsets in the family has a nonempty intersection, then \mathcal{F} can be extended to a family which is maximal with respect to having the FIP.*

Proof. Let $\tilde{\mathcal{F}}$ be the collection of all families of subsets of X that contain \mathcal{F} and satisfy the FIP. Define an order on $\tilde{\mathcal{F}}$ via inclusion and let $\{\tilde{\mathcal{F}}_j : j \in J\}$ be any chain in $\tilde{\mathcal{F}}$ – a totally ordered subset of $\tilde{\mathcal{F}}$. If $\mathcal{S} = \bigcup_{j \in J} \tilde{\mathcal{F}}_j$, then \mathcal{S} contains \mathcal{F} and if $\{s_1, \dots, s_n\}$ is any finite collection of members of \mathcal{S} then $\{s_1, \dots, s_n\} \subseteq \tilde{\mathcal{F}}_{j_0}$ for some j_0 since $\{\tilde{\mathcal{F}}_j\}_{j \in J}$ is ordered by inclusion. Now $\tilde{\mathcal{F}}_{j_0}$ enjoys the FIP. So $\bigcap_{j=1}^n s_j \neq \emptyset$; that is to say, \mathcal{S} has the FIP but also each chain in \mathcal{S} has an upper bound. From Zorn's lemma we conclude that \mathcal{S} has a maximal element, whence establishing the desired result. \square

Using this lemma, we can derive the following Andrey Nikolayevich Tychonoff's theorem.

Theorem 9.4.10. *Let $\{(X_j, \mathcal{T}_j)\}_{j \in I}$ be a family of compact topological spaces indexed by the set I . Then their product topological space $(X = \prod_{j \in I} X_j, \mathcal{T})$ is compact.*

Proof. By forming the contrapositive statement and taking complements, we immediately see that (X, \mathcal{T}) is compact if and only if any collection of closed subsets of (X, \mathcal{T}) obeying the FIP has a nonempty intersection. For this reason, we are about to make the following consideration where \mathcal{T}_j and \mathcal{T} will be dropped from (X_j, \mathcal{T}_j) and (X, \mathcal{T}) .

Suppose that \mathcal{F} is a collection of closed subsets of X which enjoys the FIP. By Lemma 9.4.9 we can extend \mathcal{F} to \mathcal{G} which is maximal with respect to having the FIP. Moreover, for each $j \in I$ the family $\{p_j(B) : B \in \mathcal{G}\}$ of subsets of X_j has the FIP since for any finite collection $\{B_1, \dots, B_n\}$ of elements in \mathcal{G} one has

$$\emptyset \neq p_j(\bigcap_{k=1}^n B_k) \subseteq \bigcap_{k=1}^n p_j(B_k) \subseteq \bigcap_{k=1}^n \overline{p_j(B_k)}.$$

Note that each X_j is compact. By the equivalent statement on the compactness there exists an x_j in $\cap_{B \in \mathcal{G}} \overline{p_j(B)}$.

Define $f \in X$ by setting $f(j) = x_j$. Since $\cap_{B \in \mathcal{G}} \overline{B} \subseteq \cap_{B \in \mathcal{F}} \overline{B}$, once verifying $f \in \cap_{B \in \mathcal{G}} \overline{B}$, we complete the argument. In so doing, let O be an arbitrary open subset of X containing f . Then for some finite set of indices j_1, \dots, j_n and open sets $O_{j_k} \subseteq X_{j_k}$, we have $f \in \cap_{k=1}^n p_{j_k}^{-1}(O_{j_k}) \subseteq O$. Accordingly, $x_{j_k} \in O_{j_k}$ and thus $O_{j_k} \cap \overline{p_{j_k}(B)} \neq \emptyset$ for all $B \in \mathcal{G}$. Using the fact that O_{j_k} is open and p_{j_k} is a projection, we can readily find $O_{j_k} \cap p_{j_k}(B) \neq \emptyset$ and hence $p_{j_k}^{-1}(O_{j_k}) \cap B \neq \emptyset$ for all $B \in \mathcal{G}$. Because \mathcal{G} is maximal with respect to enjoying the FIP, we must then have $p_{j_k}^{-1}(O_{j_k}) \in \mathcal{G}$ and similarly $\cap_{k=1}^n p_{j_k}^{-1}(O_{j_k}) \in \mathcal{G}$, whence obtaining $O \in \mathcal{G}$. From this we see that O intersects every element of \mathcal{G} and so that $f \in \overline{B}$ for all $B \in \mathcal{G}$ due to $f \in O$ and O being an arbitrary open set. \square

Now, we are ready to discuss the well-known Leonidas Alaoglu's theorem (also known as the Banach-Alaoglu theorem).

Theorem 9.4.11. *Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} . Then the closed unit ball*

$$\overline{B}_1^{X^*}(0) = \{F \in X^* : \|F\| \leq 1\}$$

of X^ is weak* compact.*

Proof. For $x \in X$ let

$$\mathbf{D}_x = \{\lambda \in \mathbf{F} : |\lambda| \leq \|x\|_X\}.$$

Note that $|F(x)| \leq \|x\|_X$ for all $F \in \overline{B}_1^{X^*}(0)$. So $\overline{B}_1^{X^*}(0)$ is a subset of the set of all functions $f : X \rightarrow \mathbf{D}_x$, i.e.,

$$\overline{B}_1^{X^*}(0) \subseteq \prod_{x \in X} \mathbf{D}_x.$$

The product topology on $\prod_{x \in X} \mathbf{D}_x$ has as open sets unions of neighborhoods of the form

$$\{f : |f(x_1) - f_0(x_1)| < \varepsilon, \dots, |f(x_n) - f_0(x_n)| < \varepsilon\}$$

for given $f_0, \varepsilon > 0$ and $x_1, \dots, x_n \in X$. These open sets when restricted to $\overline{B}_1^{X^*}(0)$ are just the weak* open subsets of $\overline{B}_1^{X^*}(0)$. According to Theorem 9.4.10, the compactness of \mathbf{D}_x implies the compactness of $\prod_{x \in X} \mathbf{D}_x$. Thus to prove that $\overline{B}_1^{X^*}(0)$ is weak* compact it is enough to show that $\overline{B}_1^{X^*}(0)$ is weak* closed since $\prod_{x \in X} \mathbf{D}_x$ is weak* compact and the argument for Theorem 6.3.7 (i) applicable.

Suppose that G is in the weak* closure of $\overline{B}_1^{X^*}(0)$; that is, each weak* neighborhood of G intersects $\overline{B}_1^{X^*}(0)$. Then we must check that G is linear and $\|G\| \leq 1$. Let $x, y \in X$ and $\varepsilon > 0$. By the assumption on G we see that the neighborhood $N(G; 3^{-1}\varepsilon, x, y, x+y)$ contains some $F \in \overline{B}_1^{X^*}(0)$. Since $F(x+y) = F(x) + F(y)$, we have

$$\begin{aligned} |G(x) + G(y) - G(x+y)| &\leq |G(x) - F(x)| + |G(y) - F(y)| + |G(x+y) - F(x+y)| \\ &< 3^{-1}\varepsilon + 3^{-1}\varepsilon + 3^{-1}\varepsilon \\ &= \varepsilon. \end{aligned}$$

So, $G(x+y) = G(x) + G(y)$. In a similar manner, we can prove $G(\lambda x) = \lambda G(x)$ for $\lambda \in \mathbf{F}$. Furthermore,

$$\begin{aligned} |F(x)| \leq \|x\|_X &\Rightarrow |G(x)| \leq |G(x) - F(x)| + |F(x)| \leq 3^{-1}\varepsilon + \|x\|_X \\ &\Rightarrow |G(x)| \leq \|x\|_X \\ &\Rightarrow \|G\| \leq 1 \\ &\Rightarrow G \in \overline{B}_1^{X^*}(0). \end{aligned}$$

Accordingly, $\overline{B}_1^{X^*}(0)$ is weak* closed and accordingly weak* compact. \square

The above-established theorem can be utilized to derive the weak compactness of the closed unit ball of a reflexive normed linear space.

Corollary 9.4.12. *Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} . If X is reflexive, then the closed unit ball of X is weakly compact.*

Proof. Recall

$$\overline{B}_1^X(0) = \{x \in X : \|x\|_X \leq 1\}.$$

Since X is reflexive, we conclude that

$$\widetilde{\overline{B}_1^X(0)} = \{\tilde{x} \in X^{**} : x \in \overline{B}_1^X(0)\}$$

is exactly the closed unit ball of X^{**} . From Theorem 9.4.11 it follows that $\widetilde{\overline{B}_1^X(0)}$ is weak* compact. But since X is reflexive, this says that $\overline{B}_1^X(0)$ is weakly compact. \square

9.5 Compact and Dual Operators

Having had a careful look at the concept of a compact (or weakly compact, or weak* compact) set, we can find that this concept is a natural generalization of a finite set. In this section we discuss the so-called compact operators and their dual forms on the general Banach spaces.

Let X and Y be two normed linear spaces over \mathbf{F} . If $T \in \mathcal{B}(X, Y)$ has a finite dimensional range $T(X)$ then the image of any bounded set in X is bounded in $T(X)$ and hence has compact closure owing to $T(X)$ being isomorphic to a Euclidean space. This motivates the following definition of a compact operator.

Definition 9.5.1. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces over \mathbf{F} . Then a linear operator $T : X \rightarrow Y$ is called compact provided that the closure $\overline{T(B)}$ is a compact subset of Y for each bounded subset B of X .*

Here we should note that Definition 9.5.1 does not require T to be continuous or bounded. Nevertheless, the following result tells us that the continuity or the boundedness of T is an immediate consequence of the definition.

Theorem 9.5.2. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces over \mathbf{F} . If a linear operator $T : X \rightarrow Y$ is compact, then it is bounded.*

Proof. In this case, $\overline{T(\overline{B_1^X})}$ is compact for the closed unit ball $\overline{B_1^X} = \{x \in X : \|x\|_X \leq 1\}$, so it is bounded in Y . Accordingly, there exists a constant $C > 0$ such that $\|y\|_Y \leq C$ for all $y \in \overline{T(\overline{B_1^X})}$. In particular, $\|T(x)\|_Y \leq C$ for $x \in \overline{B_1^X}$, namely, $\|T\| \leq C$. \square

Remark 9.5.3. *Definition 9.5.1 actually produces three more facts.*

- (i) *A linear combination of compact operators is also compact.*
- (ii) *If $T \in \mathcal{B}(X)$ is compact and $S \in \mathcal{B}(X)$ then TS and ST are compact and hence the class of all compact operators in $\mathcal{B}(X)$ form a two sided ideal in algebraic terms.*
- (iii) *A linear operator $T \in \mathcal{B}(X, Y)$ is compact if and only if for every bounded sequence $\{x_j\}_{j=1}^\infty$ in X , the sequence $\{T(x_j)\}_{j=1}^\infty$ has a convergent subsequence in Y – the necessity is obvious and to see the sufficiency, assume the statement after the if and only if; then for any bounded set $B \subseteq X$ and any sequence y_k in $\overline{T(B)}$ there is an $x_k \in B$ such that $\|y_k - T(x_k)\|_Y < k^{-1}$ and consequently $\{T(x_k)\}_{k=1}^\infty$ has and hence $\{y_k\}_{k=1}^\infty$ has a convergent subsequence with limit in $\overline{T(B)}$, establishing the compactness of $\overline{T(B)}$.*

Using the criterion in Remark 9.5.3 (iii), we obtain that the compactness is preserved under the uniform limit in the Banach space setting.

Theorem 9.5.4. *Let $(X, \|\cdot\|_X)$ be a normed linear space and $(Y, \|\cdot\|_Y)$ a Banach space over \mathbf{F} . If each $T_j \in \mathcal{B}(X, Y)$ is compact and $\lim_{j \rightarrow \infty} \|T_j - T\| = 0$ then $T : X \rightarrow Y$ is compact.*

Proof. Suppose that $\{x_j\}_{j=1}^\infty$ is a bounded sequence in X . Since T_1 is compact, there is a subsequence $\{x_{1,m}\}_{m=1}^\infty$ of $\{x_j\}_{j=1}^\infty$ such that $\{T_1(x_{1,m})\}_{m=1}^\infty$ is convergent. Also since T_2 is compact, there is a subsequence $\{x_{2,m}\}_{m=1}^\infty$ of $\{x_{1,m}\}_{m=1}^\infty$ such that $\{T_2(x_{2,m})\}_{m=1}^\infty$ is convergent. Continuing this process, we can obtain a subsequence $\{x_{n+1,m}\}_{m=1}^\infty$ of $\{x_{n,m}\}_{m=1}^\infty$ such that $\{T_{n+1}(x_{n+1,m})\}_{m=1}^\infty$ is convergent. So, $\{T_n(x_{m,m})\}_{m=1}^\infty$ is convergent for any $n \in \mathbf{N}$.

Note that $\{x_{m,m}\}_{m=1}^\infty$ is bounded in X . So $c = 1 + \sup_{m \in \mathbf{N}} \|x_{m,m}\|_X < \infty$. Given $\varepsilon > 0$, $\lim_{j \rightarrow \infty} \|T_j - T\| = 0$ implies that there is an $n_0 \in \mathbf{N}$ such that $\|T_{n_0} - T\| < \varepsilon(3c)^{-1}$. Since $\{T_{n_0}(x_{m,m})\}_{m=1}^\infty$ is convergent in Y , there is an $n_1 \in \mathbf{N}$ such that

$$k, l > n_1 \implies \|T_{n_0}(x_{k,k}) - T_{n_0}(x_{l,l})\|_Y < 3^{-1}\varepsilon.$$

With this, we achieve

$$\begin{aligned} & \|T(x_{k,k}) - T(x_{l,l})\|_Y \\ & \leq \|T(x_{k,k}) - T_{n_0}(x_{k,k})\|_Y + \|T_{n_0}(x_{k,k}) - T_{n_0}(x_{l,l})\|_Y + \|T_{n_0}(x_{l,l}) - T(x_{l,l})\|_Y \\ & \leq \|T - T_{n_0}\| \|x_{k,k}\|_X + 3^{-1}\varepsilon + \|T - T_{n_0}\| \|x_{l,l}\|_X \\ & < \varepsilon, \end{aligned}$$

and so $\{T(x_{m,m})\}_{m=1}^\infty$ is convergent in Y owing to that $(Y, \|\cdot\|_Y)$ is a Banach space. \square

Example 9.5.5. *For $1 \leq p \leq \infty$ let the linear operator T on $\ell_p(\mathbf{N}, \mathbf{F})$ be defined by $T : (x_k)_{k=1}^\infty \mapsto (k^{-1}x_k)_{k=1}^\infty$. Then T is a compact operator on $\ell_p(\mathbf{N}, \mathbf{F})$.*

To check the result, set

$$T_n : (x_k)_{k=1}^\infty \mapsto (x_1, 2^{-1}x_2, \dots, n^{-1}x_n, 0, \dots) \text{ for all } n \in \mathbf{N}.$$

It is clear that each $T_n(\ell_p(\mathbf{N}, \mathbf{F}))$ is finite dimensional and so that T_n is compact. Note that for any $x = (x_k)_{k=1}^\infty \in \ell_p(\mathbf{N}, \mathbf{F})$,

$$\|(T - T_n)((x_k)_{k=1}^\infty)\|_p = \begin{cases} \left(\sum_{k=n+1}^\infty k^{-p} |x_k|^p \right)^{\frac{1}{p}} \leq (n+1)^{-1} \|(x_k)\|_p, & p \in [1, \infty) \\ \sup_{k \geq n+1} k^{-1} |x_k| \leq (n+1)^{-1} \|(x_k)\|_\infty, & p = \infty \end{cases}.$$

So, $\lim_{n \rightarrow \infty} \|\mathsf{T} - \mathsf{T}_n\| = 0$. By Theorem 9.5.4 we can now conclude that T is a compact operator.

Next, we consider the dual operator or the adjoint or the adjoint operator of a bounded operator and see how the compactness of the original operator is reflected in the behavior of its adjoint.

Theorem 9.5.6. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces over \mathbf{F} . For $\mathsf{T} \in \mathcal{B}(X, Y)$ set $\mathsf{T}^*(f) = f \circ \mathsf{T}$. Then $\mathsf{T}^* \in \mathcal{B}(Y^*, X^*)$ and $\|\mathsf{T}^*\| = \|\mathsf{T}\|$. Furthermore, if X and Y are Banach spaces, then T is compact when and only when T^* is compact.*

Proof. If $f \in Y^*$ and $x \in X$, it is easy to see that $\mathsf{T}^*(f)(x) = f \circ \mathsf{T}(x)$ defines an \mathbf{F} -valued linear function on X since f and T are linear. Consequently,

$$|\mathsf{T}^*(f)(x)| \leq \|f\| \|\mathsf{T}(x)\|_Y \leq \|f\| \|\mathsf{T}\| \|x\|_X;$$

that is, $\mathsf{T}^*(f) \in X^*$. Clearly, $\mathsf{T}^* : Y^* \rightarrow X^*$ is linear and bounded. According to Corollary 8.4.4 (vi), we evaluate

$$\begin{aligned} \|\mathsf{T}\| &= \sup_{\|x\|_X=1} \|\mathsf{T}(x)\|_Y \\ &= \sup_{\|x\|_X=1} \sup_{\|f\|=1} |f \circ \mathsf{T}(x)| \\ &= \sup_{\|f\|=1} \sup_{\|x\|_X=1} |\mathsf{T}^*(f)(x)| \\ &= \sup_{\|f\|=1} \|\mathsf{T}^*(f)\| \\ &= \|\mathsf{T}^*\|. \end{aligned}$$

To check the second part, it is enough to verify that if T is compact then so is T^* due to the symmetry between both operators. Suppose that T is compact. Let $\{f_j\}_{j=1}^\infty$ be a sequence in the closed unit ball of Y^* , i.e., $\sup_{j \in \mathbf{N}} \|f_j\| \leq 1$. Then

$$|f_j(y_1) - f_j(y_2)| \leq \|y_1 - y_2\|_Y \quad \text{for all } y_1, y_2 \in Y,$$

and hence $\{f_j\}_{j=1}^\infty$ is equi-continuous. Since the closure $\overline{\mathsf{T}(\overline{B})}$ of the image $\mathsf{T}(\overline{B}) \subseteq Y$ of the closed unit $\overline{B} \subseteq X$ is compact, an application of Theorem 7.3.7 yields that $\{f_j\}_{j=1}^\infty$ has a subsequence $\{f_{j_k}\}_{k=1}^\infty$ which converges uniformly on $\mathsf{T}(\overline{B})$. Because of

$$\|\mathsf{T}^*(f_{j_k}) - \mathsf{T}^*(f_{j_l})\| = \sup_{\|x\|_X=1} |f_{j_k} \circ \mathsf{T}(x) - f_{j_l} \circ \mathsf{T}(x)|,$$

the completeness of X^* implies that $\{\mathbb{T}^*(f_{j_k})\}_{k=1}^\infty$ converges in X^* . Accordingly, \mathbb{T}^* is compact. \square

Example 9.5.7. Let $[a, b]$ and $[c, d]$ be two finite closed intervals of \mathbf{R} . For $k(\cdot, \cdot) \in C([c, d] \times [a, b], \mathbf{F})$ let $\mathbb{T} : C([a, b], \mathbf{F}) \rightarrow C([c, d], \mathbf{F})$ be given by

$$\mathbb{T}(f)(x) = \int_a^b k(x, y)f(y)dy \text{ for all } x \in [c, d].$$

Then \mathbb{T} is a compact operator and so is its adjoint \mathbb{T}^* which maps

$$C([c, d], \mathbf{F})^* = BV([c, d], \mathbf{F})/BV_0([c, d], \mathbf{F})$$

into

$$C([a, b], \mathbf{F})^* = BV([a, b], \mathbf{F})/BV_0([a, b], \mathbf{F})$$

by Theorem 9.5.6, where

$$\mathbb{T}^*(g)(y) = \int_c^d \overline{k(x, y)}dg(x) \text{ for all } y \in [a, b],$$

which can be determined by Theorem 9.3.4.

To verify the compactness of \mathbb{T} , we notice that $k(\cdot, \cdot) \in C([c, d] \times [a, b], \mathbf{F})$ implies

$$\mathbf{M}_1 = \|k(\cdot, \cdot)\|_\infty = \sup_{(x, y) \in [c, d] \times [a, b]} |k(x, y)| < \infty,$$

and thus get that if B is a bounded subset of $C([a, b], \mathbf{F})$; that is, there is a constant $\mathbf{M}_2 > 0$ such that $\|f\|_\infty \leq \mathbf{M}_2$ for all $f \in B$, then

$$\|\mathbb{T}(f)\|_\infty = \sup_{x \in [c, d]} \left| \int_a^b k(x, y)f(y)dy \right| \leq \mathbf{M}_1 \mathbf{M}_2 (b - a).$$

This means that $\mathbb{T}(B)$ is bounded in $C([c, d], \mathbf{F})$. Since $k(\cdot, \cdot)$ is indeed uniformly continuous on $[c, d] \times [a, b]$, we see that for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$x_1, x_2 \in [c, d] \text{ and } |x_1 - x_2| < \delta \Rightarrow |k(x_1, y) - k(x_2, y)| < \varepsilon \mathbf{M}_2^{-1} (b - a)^{-1}.$$

From this implication we further see that

$$x_1, x_2 \in [c, d] \text{ and } |x_1 - x_2| < \delta \Rightarrow |\mathbb{T}(f)(x_1) - \mathbb{T}(f)(x_2)| \leq \varepsilon$$

for all $f \in B$. By Theorem 7.3.7 we can now conclude that the closure $\overline{\mathbb{T}(B)}$ is compact in $C([c, d], \mathbf{F})$ and so that \mathbb{T} is a compact operator.

Problems

9.1. Prove that if $C^n[0, 1]$ is the class of all functions $f : [0, 1] \rightarrow \mathbf{R}$ for which the derivatives $f^{(0)}, f^{(1)}, \dots, f^{(n)}$ belong to $C[0, 1]$ and satisfy $\|f\| = \sup_{0 \leq k \leq n} \sup_{t \in [0, 1]} |f^{(k)}(t)| < \infty$, then it is a Banach space under this norm $\|\cdot\|$.

9.2. Prove that if $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are Banach spaces over \mathbf{F} then so is $X \times Y$ under the norm $\|(x, y)\| = \max\{\|x\|_X, \|y\|_Y\}$ for $(x, y) \in X \times Y$.

9.3. Prove that $(C[0, 1], \|\cdot\|_p)$ with $p \in [1, \infty)$ is not complete.

9.4. Let $L^1(\mathbf{Z}, \mathbf{F})$ comprise all \mathbf{F} -valued functions f on \mathbf{Z} obeying $\|f\|_1 = \sum_{n=-\infty}^{\infty} |f(n)| < \infty$. Prove that this space is a Banach algebra under the following convolution

$$f * g(n) = \sum_{m=-\infty}^{\infty} f(n-m)g(m) \text{ for all } f, g \in L^1(\mathbf{Z}, \mathbf{F}).$$

9.5. Show by an example that the uniform bounded principle does not hold once the completeness is dropped.

9.6. Prove that there is a continuous map $f : \mathbf{R} \rightarrow \mathbf{R}$ such that $f(U)$ is not open even provided that U is open.

9.7.

(i) Prove

$$\lim_{n \rightarrow \infty} \int_0^{2\pi} |\sin(n + 2^{-1})x| |\sin 2^{-1}x|^{-1} dx = \infty.$$

(ii) Given a Riemann-integrable function $f : [0, 2\pi] \rightarrow \mathbf{R}$, let its Fourier series be

$$s(x) = \sum_{m=-\infty}^{\infty} a_m e^{imx} \text{ where } a_m = (2\pi)^{-1} \int_0^{2\pi} f(y) e^{-imy} dy.$$

Extend the definition of f to make it 2π -periodic. Define the n -th partial sum of the Fourier series to be $s_n(x) = \sum_{m=-n}^n a_m e^{imx}$. Prove

$$s_n(x) = (2\pi)^{-1} \int_0^{2\pi} f(x+y) \left(\frac{\sin(n+2^{-1})y}{\sin 2^{-1}y} \right) dy \text{ for all } x \in (0, 2\pi).$$

(iii) Let X be the Banach space of continuous functions $f : [0, 2\pi] \rightarrow \mathbf{R}$ with $f(0) = f(2\pi)$, with the sup-norm. Prove that the linear operator $T_n : X \rightarrow \mathbf{R}$ defined by

$$T_n(f) = (2\pi)^{-1} \int_0^{2\pi} f(x) \left(\frac{\sin(n+2^{-1})x}{\sin 2^{-1}x} \right) dx$$

is bounded, and

$$\|T_n\| = (2\pi)^{-1} \int_0^{2\pi} \left| \frac{\sin(n+2^{-1})x}{\sin 2^{-1}x} \right| dx.$$

(iv) Prove that there exists a continuous function $f : [0, 2\pi] \rightarrow \mathbf{R}$ with $f(0) = f(2\pi)$ such that its Fourier series diverges at $x = 0$.

9.8. Let $X = C^1[0, 1]$ and $Y = C[0, 1]$ be equipped with the sup-norm. Prove the following results.

(i) X is not complete.

(ii) The map $(d/dx) : X \rightarrow Y$ is closed but not bounded.

9.9. Suppose that $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are Banach spaces over \mathbf{F} and $T \in \mathcal{B}(X, Y)$ satisfying $\|T(x)\|_Y \geq c\|x\|_X$ for all $x \in X$ and for some constant $c > 0$. Verify the following facts:

(i) $N(T) = \{x \in X : T(x) = 0\} = \{0\}$;

(ii) $R(T) = \{y \in Y : y = T(x) \text{ for some } x \in X\}$ is closed;

(iii) $T^{-1} \in \mathcal{B}(R(T), X)$ satisfies $\|T^{-1}\| \leq c^{-1}$.

9.10. Prove the following two results:

(i) $\{f_j\}_{j=1}^\infty$ converges weakly to f in the real space $C[0, 1]$ if and only if $\sup_{j \in \mathbf{N}} \|f_j\|_\infty < \infty$ and $\lim_{j \rightarrow \infty} f_j(x) = f(x)$ for any $x \in [0, 1]$.

(ii) $\{x_j = (s_k^{(j)})_{k=1}^\infty\}_{j=1}^\infty$ converges weakly to $x = (s_k)_{k=1}^\infty$ in the real sequence space $\ell_p = \ell_p(\mathbf{N}, \mathbf{R})$ where $p \in (1, \infty)$, if and only if, $\sup_{j \in \mathbf{N}} \|x_j\|_p < \infty$ and $\lim_{j \rightarrow \infty} s_k^{(j)} = s_k$ for any $k \in \mathbf{N}$.

9.11. Let $(X, \|\cdot\|_X)$ be a separable Banach space over \mathbf{F} and M be a bounded subset of X^* . Prove that every sequence in M contains a weak* convergent subsequence.

9.12. Let $(X, \|\cdot\|_X)$ be a Banach space over \mathbf{F} . Show that X is reflexive if and only if X^* is reflexive.

9.13. Let $c_0 = c_0(\mathbf{N}, \mathbf{R})$, the linear space consisting of all $\ell_\infty = \ell_\infty(\mathbf{N}, \mathbf{R})$ -sequences that converge to 0. Prove the following two facts:

- (i) c_0 is a separable, but non-reflexive Banach space over \mathbf{R} ;
- (ii) $L \in \mathcal{B}(c_0, \mathbf{R})$ if and only if there is an element $y = (y_k)_{k=1}^\infty \in \ell_1 = \ell_1(\mathbf{N}, \mathbf{R})$ such that $L(x) = \sum_{k=1}^\infty x_k y_k$ for all $x = (x_k)_{k=1}^\infty \in c_0$.

9.14. Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} and K be a convex subset of X . Prove that the weak closure of K equals its norm closure.

9.15. Let $(X, \|\cdot\|_X)$ be a normed linear space over \mathbf{F} and $A \subseteq X^*$ be weak* compact. Prove that:

- (i) A is weak* closed;
- (ii) any weak* closed subset of A is weak* compact.

9.16. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces over \mathbf{F} . Prove that if $\mathsf{T} : X \rightarrow Y$ is a linear mapping such that $f \circ \mathsf{T} \in X^*$ for all $f \in Y^*$ then $\mathsf{T} \in \mathcal{B}(X, Y)$.

9.17. Prove that a topological space X is compact if and only if any collection of closed subsets of X obeying the FIP has a nonempty intersection.

9.18. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear and Banach spaces over \mathbf{F} respectively. Prove that the set $\mathcal{B}_c(X, Y)$ of all compact operators from X to Y is a Banach space under the usual operator norm for $\mathcal{B}(X, Y)$.

9.19. Given a finite interval $[a, b] \subset \mathbf{R}$, let $k(\cdot, \cdot)$ be an \mathbf{F} -valued continuous function on $\{(x, y) \in \mathbf{R}^2 : x \in [a, b], y \in [a, x]\}$ and let $\mathsf{T} : C([a, b], \mathbf{F}) \rightarrow C([a, b], \mathbf{F})$ be given by

$$\mathsf{T}(f)(x) = \int_a^x k(x, y)f(y)dy.$$

Show that T is a compact operator.

9.20. Suppose that X is a Banach space over \mathbf{F} and $\mathsf{T} \in \mathcal{B}(X)$ is a compact operator. Prove the following assertions:

- (i) TS and ST are compact for any $\mathsf{S} \in \mathcal{B}(X)$;
- (ii) $\mathsf{T}(X)$ is separable.

9.21. Let $(X, \|\cdot\|_X)$ be a Banach space over \mathbf{F} . Prove the following three results.

(i) If $T : X \rightarrow X$ is the zero operator; that is, $T(x) = 0$ for all $x \in X$, then T is compact;

(ii) If $T \in \mathcal{B}(X)$ is compact and X is infinite dimensional, then T is not invertible;

(iii) If $T \in \mathcal{B}(X)$ ensures that $T(X)$ is closed and infinite dimensional, then T is not compact.

Chapter 10

Hilbert Spaces and Their Operators

Hilbert spaces, named after David Hilbert, are special Banach spaces which share most of the rich geometric structure of the finite-dimensional Euclidean spaces. The geometrical wealth of a Hilbert space is produced by a bilinear form on the space which allows the introduction of a notion of perpendicularity. This chapter is devoted to an introductory study of the Hilbert spaces and their bounded and compact operators.

10.1 Definition, Examples and Basic Properties

From now on, \mathbf{F} still denotes the field of real or complex numbers.

Definition 10.1.1. *Let X be a linear space over \mathbf{F} . An inner product on X is a map $(x, y) \mapsto \langle x, y \rangle$ from $X \times X$ to \mathbf{F} such that:*

- (a) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$ for all $x, y, z \in X$ and for all $\alpha, \beta \in \mathbf{F}$;
- (b) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in X$;
- (c) $\langle x, x \rangle \geq 0$ for all $x \in X$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

If X is equipped with $\langle \cdot, \cdot \rangle$, then it is called an inner product space or a pre-Hilbert space over \mathbf{F} .

Note that the bar signifies complex conjugation – of course, if the scalar field is \mathbf{R} then the axiom (b) is simply $\langle y, x \rangle = \langle x, y \rangle$. So we will refer to real or complex Hilbert spaces depending on whether \mathbf{F} is \mathbf{R} or \mathbf{C} .

Theorem 10.1.2. *Let X be a linear space over \mathbf{F} under the inner product $\langle \cdot, \cdot \rangle$. If $\|x\| = \sqrt{\langle x, x \rangle}$ for all $x \in X$, then there are the following statements.*

(i) *Cauchy-Schwarz's inequality $|\langle x, y \rangle| \leq \|x\| \|y\|$ holds for all $x, y \in X$, with equality if and only if x and y are linearly dependent;.*

(ii) *$\|\cdot\|$ defines a norm on X .*

(iii) *The parallelogram law $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$ is valid for all $x, y \in X$.*

Proof. (i) If $\langle x, y \rangle = 0$, then there is nothing to argue. Suppose that $\langle x, y \rangle \neq 0$. Then $x \neq 0$ and $y \neq 0$. Let $\alpha = \langle x, y \rangle$ and $z = \alpha y$. Then

$$0 \leq \langle x - tz, x - tz \rangle = \|x\|^2 - 2t|\langle x, y \rangle|^2 + t^2|\langle x, y \rangle|^2 \|y\|^2 \quad \text{for all } t \in \mathbf{R}.$$

This ensures that the last quadratic function of t takes its absolute minimum at $t = \|y\|^{-2}$. Substituting this value for t , we get

$$0 \leq \|x - tz\|^2 = \|x\|^2 - \|y\|^{-2} |\langle x, y \rangle|^2$$

with equality if and only if $x - tz = x - \alpha y = 0$, from which the desired result is immediate.

(ii) It is obvious that $\|x\| = 0 \Leftrightarrow x = 0$ and $\|\lambda x\| = |\lambda| \|x\|$. As for the triangle inequality, (i) is used to imply

$$\|x+y\|^2 = \|x\|^2 + 2\Re \langle x, y \rangle + \|y\|^2 \leq (\|x\| + \|y\|)^2.$$

(iii) This follows directly from expanding $\|x+y\|^2$ and $\|x-y\|^2$. □

The above definition and theorem actually induce the notion of a Hilbert space.

Definition 10.1.3. *A Hilbert space X over \mathbf{F} is a pre-Hilbert space which is complete with respect to the metric*

$$d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y \rangle} \quad \text{for all } x, y \in X.$$

Furthermore, we will refer to real or complex Hilbert space according to whether \mathbf{F} is \mathbf{R} or \mathbf{C} .

Example 10.1.4.

(i) For each $n \in \mathbf{N}$, \mathbf{F}^n is a Hilbert space over \mathbf{F} under the inner product

$$\langle x, y \rangle = \sum_{j=1}^n x_j \overline{y_j} \text{ for all } x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbf{F}^n.$$

(ii) Regardless of the cardinality of an arbitrarily given set X , we can define the sequence space

$$\ell_2(X, \mathbf{F}) = \left\{ f : X \rightarrow \mathbf{F} : \sum_{x \in X} |f(x)|^2 < \infty \right\},$$

where

$$\sum_{x \in X} |f(x)|^2 = \sup \left\{ \sum_{x \in Y} |f(x)|^2 : \text{all nonempty finite subsets } Y \text{ of } X \right\}.$$

This space becomes a Hilbert space over \mathbf{F} with the inner product

$$\langle f_1, f_2 \rangle = \sum_{x \in X} f_1(x) \overline{f_2(x)} \text{ for all } f_1, f_2 \in \ell_2(X, \mathbf{F}).$$

Here the right-hand sum is said to converge provided there is a number s in \mathbf{F} such that to each $\epsilon > 0$ there corresponds a finite subset Y of X with the property that if Z is any finite subset of X with $Y \subseteq Z$ then

$$\left| \sum_{x \in Z} f_1(x) \overline{f_2(x)} - s \right| < \epsilon.$$

Obviously, $\ell_2(\mathbf{N}, \mathbf{F})$ is just the square-summable \mathbf{F} -valued sequence space.

(iii) $LRS_{id}^2([a, b], \mathbf{F})$, often denoted $L^2([a, b], \mathbf{F})$, is a Hilbert space over \mathbf{F} under the inner product

$$\langle f_1, f_2 \rangle = \int_{[a, b]} f_1 \overline{f_2} dm_{id} \text{ for all } f_1, f_2 \in LRS_{id}^2([a, b], \mathbf{F}).$$

The following result reveals when a Banach space becomes a Hilbert space.

Theorem 10.1.5. Let $(X, \|\cdot\|)$ be a normed linear space over \mathbf{F} and satisfy the parallelogram law.

(i) If $\mathbf{F} = \mathbf{C}$, then

$$\langle x, y \rangle = 2^{-2} \sum_{n=1}^4 i^n \|x + i^n y\|^2$$

defines an inner product on X .

(ii) If $\mathbf{F} = \mathbf{R}$, then

$$\langle x, y \rangle = 2^{-2} (\|x + y\|^2 - \|x - y\|^2)$$

defines an inner product on X .

Proof. It is enough to check (i). As a matter of fact, this $\langle \cdot, \cdot \rangle$ enjoys

$$\langle x, x \rangle = \|x\|^2 + 2^{-2}i|1+i|^24\|x\|^2 - 2^{-2}i|1-i|^2\|x\|^2 = \|x\|^2 \text{ for all } x \in X.$$

To verify that $\langle \cdot, \cdot \rangle$ is actually an inner product, it suffices to prove that Definition 10.1.1 (a) holds for $\alpha = \beta = 1$ and $\langle \lambda \cdot, \cdot \rangle = \lambda \langle \cdot, \cdot \rangle$ for any $\lambda \in \mathbf{C}$.

For the former, we use the parallelogram law to achieve

$$\begin{cases} \|u + v + w\|^2 + \|u + v - w\|^2 = 2\|u + v\|^2 + 2\|w\|^2; \\ \|u - v + w\|^2 + \|u - v - w\|^2 = 2\|u - v\|^2 + 2\|w\|^2. \end{cases}$$

Hence

$$2(\|u + v\|^2 - \|u - v\|^2) = (\|u + v + w\|^2 - \|u - v + w\|^2) + (\|u + v - w\|^2 - \|u - v - w\|^2).$$

This infers

$$\Re \langle u + w, v \rangle + \Re \langle u - w, v \rangle = 2\Re \langle u, v \rangle.$$

In a similar fashion, we can establish the relation with the real part \Re replaced by the imaginary part \Im :

$$\Im \langle u + w, v \rangle + \Im \langle u - w, v \rangle = 2\Im \langle u, v \rangle.$$

Accordingly,

$$\langle u + w, v \rangle + \langle u - w, v \rangle = 2\langle u, v \rangle.$$

When $u = w$, we have $\langle 2u, v \rangle = 2\langle u, v \rangle$. Taking $u + w = x, u - w = y, v = z$, we get

$$\langle x, z \rangle + \langle y, z \rangle = 2\langle 2^{-1}(x + y), z \rangle = \langle x + y, z \rangle.$$

To reach the latter, we observe that for any $m \in \mathbf{N}$,

$$\langle mx, y \rangle = \langle (m-1)x + x, y \rangle = \langle (m-1)x, y \rangle + \langle x, y \rangle = \cdots = m\langle x, y \rangle.$$

Thus for any $k \in \mathbf{N}$,

$$k\langle k^{-1}x, y \rangle = \langle x, y \rangle \text{ and } \langle k^{-1}x, y \rangle = k^{-1}\langle x, y \rangle.$$

Consequently, for any $r = k^{-1}m$,

$$r\langle x, y \rangle = m\langle k^{-1}x, y \rangle = \langle mk^{-1}x, y \rangle = \langle rx, y \rangle.$$

Since $\lim_{\|x\| \rightarrow 0} \|x + i^n y\| = \|y\|$, $\langle x, y \rangle$ is a continuous functional in x , and consequently $\lambda\langle x, y \rangle = \langle \lambda x, y \rangle$ for any $\lambda > 0$ thanks to the density of \mathbf{Q} in \mathbf{R} . If $\lambda < 0$ then

$$\lambda\langle x, y \rangle - \langle \lambda x, y \rangle = \lambda\langle x, y \rangle - |\lambda|\langle -x, y \rangle = \lambda\langle 0, y \rangle = 0.$$

Also, it is evident to verify

$$i\langle x, y \rangle = 2^{-2} \sum_{n=1}^4 \|x + i^{n-1}y\| = \langle ix, y \rangle.$$

Therefore, for any $\lambda = \mu + i\nu \in \mathbf{C}$ we have

$$\lambda\langle x, y \rangle = \mu\langle x, y \rangle + i\langle \nu x, y \rangle = \langle (\mu + i\nu)x, y \rangle,$$

as desired. □

10.2 Orthogonality, Orthogonal Complement and Duality

First of all, let us consider orthogonality.

Definition 10.2.1. *Let X be a Hilbert space over \mathbf{F} .*

(i) *The angle between two vectors $x, y \in X$ is defined by*

$$\theta_{x,y} = \begin{cases} 0 & , x \text{ or } y = 0 \\ \arccos \frac{\Re\langle x, y \rangle}{\|x\|\|y\|} & , \text{ otherwise } \end{cases}.$$

(ii) *Two vectors $x, y \in X$ are called orthogonal, denoted $x \perp y$, provided $\langle x, y \rangle = 0$.*

(iii) *For any subset S of X , the set $S^\perp = \{x \in X : \langle x, y \rangle = 0 \text{ for all } y \in S\}$ is called the orthogonal complement of S .*

The following result, which has a root in Corollary 9.3.6 and will be used later, has natural geometrical and finite-dimensional antecedents.

Theorem 10.2.2. *Let X be a Hilbert space over \mathbf{F} . If M is a closed convex subset of X , then there exists exactly one $x \in M$ such that $\|x\| = \inf_{y \in M} \|y\|$.*

Proof. Let $\delta = \inf_{y \in M} \|y\|$. Then by definition there is a sequence $\{x_j\}_{j=1}^\infty$ in M such that $\lim_{j \rightarrow \infty} \|x_j\| = \delta$. Since M is convex, we conclude $2^{-1}(x_j + x_k) \in M$ and consequently,

$$\|x_k + x_j\|^2 = 4\|2^{-1}(x_k + x_j)\|^2 \geq 4\delta^2.$$

Using the parallelogram law, we obtain

$$4\delta^2 + \|x_k - x_j\|^2 \leq \|x_k + x_j\|^2 + \|x_k - x_j\|^2 = 2(\|x_k\|^2 + \|x_j\|^2).$$

Letting $j, k \rightarrow \infty$ in the last inequalities, we find that $\|x_j - x_k\| \rightarrow 0$ as $j, k \rightarrow \infty$; that is, $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence and so there is a point $x \in M$ such that $x_j \rightarrow x$ in X since M is closed. The continuity of the norm leads to $\|x\| = \delta$. Finally, x is unique since if $y \in M$ satisfies $\|y\| = \delta$ then the forgoing argument can be used for the sequence x, y, x, y, x, y, \dots to show that this sequence is a Cauchy sequence, and hence $x = y$. \square

Next, we introduce the concept of a direct sum which has implicitly appeared in Corollary 8.4.4.

Definition 10.2.3. *Let X and Y be two linear subspaces of a given linear space Z over \mathbf{F} . Then Z is said to be the direct sum of X and Y , denoted $Z = X \oplus Y$, provided that every $z \in Z$ can be expressed uniquely in the form $z = x + y$ where $x \in X$ and $y \in Y$ and $X \cap Y = \{0\}$.*

We are about to prove a theorem on decomposing a Hilbert space into a direct sum of mutually orthogonal closed subspaces. Before doing so, we need the following result.

Lemma 10.2.4. *Let M be a proper closed linear subspace of a Hilbert space X over \mathbf{F} . Then there exists a unique nonzero $z \in X$ such that $\langle z, y \rangle = 0$ for all $y \in M$.*

Proof. Given any $x \in X$, the set $x + M$ is a closed convex set. Thus by Theorem 10.2.2 there exists a unique $z \in x + M$ such that $\|z\| = \inf_{y \in M} \|x + y\|$. Since $M \neq X$, we may assume $x \notin M$ which yields $z \neq 0$, and then prove that $\langle z, y \rangle = 0$ for all $y \in M$. Given any $y \in M$ and any $\alpha \in \mathbf{F}$, we have $z + \alpha y \in x + M$. By our choice of z we then have $\|z + \alpha y\|^2 \geq \|z\|^2$ and consequently,

$$|\alpha|^2 \|y\|^2 + 2\Re(\bar{\alpha}\langle z, y \rangle) \geq 0 \text{ for all } \alpha \in \mathbf{F}.$$

Choose $\theta \in [0, 2\pi)$ with $e^{-i\theta}\langle z, y \rangle = |\langle z, y \rangle|$ and let $\alpha = -te^{i\theta}$, $t > 0$. The inequality obtained by substituting for this α is then $t\|y\|^2 \geq 2|\langle z, y \rangle|$. Letting $t \rightarrow 0$ we get $\langle z, y \rangle = 0$. Because $y \in M$ was arbitrary, we must have $\langle z, y \rangle = 0$ for all $y \in M$. \square

Theorem 10.2.5. *If M is a proper closed linear subspace of a Hilbert space X over \mathbf{F} , then $X = M \oplus M^\perp$ and M^\perp is closed.*

Proof. Given $x \in X$, apply the procedure in the proof of Lemma 10.2.4 to obtain $z \in X$ such that $z \in x + M$ and $\langle z, y \rangle = 0$ for all $y \in M$. Then $z \in M^\perp$ and $z = x - y$ for some $y \in M$. Hence $x = y + z$, $y \in M$, and $z \in M^\perp$. Noting that $M \cap M^\perp = \{0\}$ since

$$x \in M \cap M^\perp \Rightarrow \langle x, x \rangle = 0 \Rightarrow x = 0,$$

we thus have $X = M \oplus M^\perp$. To see that M^\perp is closed, suppose that $x_j \in M^\perp$ and $x_j \rightarrow x$ in X . Then for any $y \in M$ we have by the Cauchy-Schwarz inequality,

$$|\langle y, x \rangle| \leq |\langle y, x_j - x \rangle| + |\langle y, x_j \rangle| \leq \|y\| \|x_j - x\| \rightarrow 0 \text{ as } j \rightarrow \infty,$$

whence deriving $x \in M^\perp$. \square

Finally, we consider the dual space of a Hilbert space. Below is the classical Riesz representation theorem.

Theorem 10.2.6. *Given a Hilbert space X over \mathbf{F} and $y \in X$, define $L_y : X \rightarrow \mathbf{F}$ by $L_y(x) = \langle x, y \rangle$. Then $L_y \in X^*$ and $\|L_y\| = \|y\|$. Conversely, for every $f \in X^*$ there exists a unique $y \in X$ such that $f = L_y$.*

Proof. Clearly, $L_y \in X^*$ with $\|L_y\| \leq \|y\|$. Note that

$$\|y\|^2 = L_y(y) \leq \|L_y\| \|y\|.$$

So $\|L_y\| = \|y\|$. Conversely, let $f \in X^*$ be given. If $f = 0$, then $f = L_y$ where $y = 0$. Suppose that $f \neq 0$. Then we may assume without loss of generality that $\|f\| = 1$ since

$$f\|f\|^{-1} = L_y \Rightarrow f = L_{\|f\|y}.$$

For such an f , let $M = \{x \in X : f(x) = 0\}$. Then M is closed linear subspace of X . Of course, $M \neq X$ otherwise $f = 0$ – a contradiction. By Lemma 10.2.4 we can obtain a nonzero $z \in M^\perp$ such that $f(z) \neq 0$. Now for any $x \in X$, we have

$$x - f(x)(f(z))^{-1}z \in M \quad \text{and} \quad \langle x - f(x)(f(z))^{-1}z, z \rangle = 0.$$

Accordingly,

$$\langle x, z \rangle = f(x) \langle z(f(z))^{-1}, z \rangle.$$

Taking $y = \overline{f(z)}\|z\|^{-2}z$, we further get $f(x) = \langle x, y \rangle$, and then $f = L_y$. To see the uniqueness, assume there is another point $w \in X$ such that $f = L_w$. Then $\langle x, w - y \rangle = 0$ for all $x \in X$, and hence $\|w - y\|^2 = \langle w - y, w - y \rangle = 0$ and $w = y$. \square

Here is an easy consequence of Theorem 10.2.6 which says that any Hilbert space is self-reflexive.

Corollary 10.2.7. *Let X be a Hilbert space over \mathbf{F} . Then the mapping $L : X \rightarrow X^*$ given by $L(x)(\cdot) = \langle \cdot, x \rangle$ is an isometric embedding from X onto X^* . Moreover, X^* is also a Hilbert space over \mathbf{F} .*

Proof. The first part has just been proved in Theorem 10.2.6. Furthermore, it is easily seen that $L(x_1 + x_2) = L(x_1) + L(x_2)$ so the isometry is additive. Since

$$L(\alpha x)(y) = \langle y, \alpha x \rangle = \bar{\alpha} \langle y, x \rangle = \bar{\alpha} L(x)(y) \quad \text{for all } x, y \in X \text{ and } \alpha \in \mathbf{F},$$

we conclude $L(\alpha x) = \bar{\alpha} L(x)$. If X^* is now equipped with the inner product $\langle L(x), L(y) \rangle = \langle y, x \rangle$, then X^* is a Hilbert space over \mathbf{F} . \square

10.3 Orthonormal Sets and Bases

To better understand the structure of a Hilbert space, in this section we consider the concepts of orthonormal sets and bases.

Definition 10.3.1. Let X be a Hilbert space over \mathbf{F} . A set $S \subseteq X$ is called *orthogonal* if any two different elements in S are orthogonal. An *orthonormal set* is an orthogonal set consisting entirely of elements of norm 1.

A constructive method of orthonormalizing a set of vectors in a Hilbert space is the Gram-Schmidt process, named for Jørgen Pedersen Gram and Erhard Schmidt, as follows.

Theorem 10.3.2. Let $\{x_j\}_{j=1}^{\infty}$ be a countable linearly independent set of a Hilbert space X over \mathbf{F} . Then a countable orthonormal set $\{e_j\}$ can be constructed so that

$$\text{span}\{e_j\}_{j=1}^n = \text{span}\{x_j\}_{j=1}^n \text{ for all } n \in \mathbf{N}.$$

Proof. Define $e_1 = \|x_1\|^{-1}x_1$ and proceed inductively. Suppose that e_1, e_2, \dots, e_n are successively defined for each $n \in \mathbf{N}$. Then e_{n+1} is determined by $y_{n+1} \|y_{n+1}\|^{-1}$, where

$$y_{n+1} = x_{n+1} - \sum_{j=1}^n \langle x_{n+1}, e_j \rangle e_j.$$

Thus $\|y_{n+1}\| \neq 0$ since otherwise we would have $x_{n+1} \in \text{span}\{e_j\}_{j=1}^n = \text{span}\{x_j\}_{j=1}^n$, contradicting the linear independence of $\{x_j\}_{j=1}^{\infty}$. It is clear that $\text{span}\{e_j\}_{j=1}^{n+1} = \text{span}\{x_j\}_{j=1}^{n+1}$ because this is true for $\{e_j\}_{j=1}^n$ and $\{x_j\}_{j=1}^n$. Finally, for $j \in \{1, 2, \dots, n\}$, we have

$$\langle e_{n+1}, e_j \rangle = (\langle x_{n+1}, e_j \rangle - \langle x_{n+1}, e_j \rangle \langle e_j, e_j \rangle) \|y_{n+1}\|^{-1} = 0,$$

and consequently, the set $\{e_j\}_{j=1}^{\infty}$ obtained by this inductive construction is an orthonormal set with the desired property. \square

Example 10.3.3. Given $-\infty \leq a < b \leq \infty$ and a function $\omega: (a, b) \rightarrow (0, \infty)$ with the property that $\int_{(a,b)} t^n \omega(t) dm_{id}(t)$ converges for all $n \in \mathbf{N}$, define the Hilbert space $L^{2,\omega}((a,b), \mathbf{F})$ to be the linear space of all \mathbf{F} -valued functions f on (a,b) with

$$\int_{(a,b)} |f|^2 \omega dm_{id} < \infty.$$

Of course, that two functions in the space coincide stands for $f_1 = f_2$ ωdm_{id} -a.e. on (a,b) . It is not hard to prove that the linearly independent set $\{1, t, t^2, \dots\}$

has a linear span which is dense in $L^{2,\omega}((a,b),\mathbf{F})$. This set, together with the Gram-Schmidt process through the inner product

$$\langle f_1, f_2 \rangle_\omega = \int_{(a,b)} f_1 \overline{f_2} \omega dm_{id},$$

derives various families of classical orthonormal functions below.

(i) If $\omega = 1$ and $a = -1$, $b = 1$, then the process generates the Legendre polynomials, named after Adrien-Marie Legendre.

(ii) If $\omega(t) = (1-t^2)^{-\frac{1}{2}}$ and $a = -1$, $b = 1$, then the process produces the Chebychev polynomials, named after Pafnuty Chebyshev.

(iii) If $\omega(t) = t^{q-1}(1-t)^{p-q}$ with $q > 0$, $p-q > -1$ and $a = 0$, $b = 1$, then the process generates the Jacobi polynomials, named after Karl Gustav Jacob Jacobi.

(iv) If $\omega(t) = e^{-t^2}$ and $a = -\infty$, $b = \infty$, then the process produces the Hermite polynomials, named after Charles Hermite.

(v) If $\omega(t) = e^{-t}$ and $a = 0$, $b = \infty$, then the process gives the Laguerre polynomials, named after Edmond Laguerre.

Lemma 10.3.4. For $n \in \mathbf{N}$ let $\{e_j\}_{j=1}^n$ be an orthonormal set of a Hilbert space X over \mathbf{F} . Then:

- (i) $\sum_{j=1}^n |\langle x, e_j \rangle|^2 \leq \|x\|^2$ for all $x \in X$;
- (ii) $\langle x - \sum_{j=1}^n \langle x, e_j \rangle e_j, e_k \rangle = 0$ for all $x \in X$ and $k = 1, 2, \dots, n$.

Proof. (i) This follows from

$$0 \leq \left\| x - \sum_{j=1}^n \langle x, e_j \rangle e_j \right\|^2 = \|x\|^2 - \sum_{j=1}^n |\langle x, e_j \rangle|^2,$$

where the last equality is obtained by expanding in the usual fashion and using the orthonormality of $\{e_j\}_{j=1}^n$.

(ii) This follows from

$$\left\langle x - \sum_{j=1}^n \langle x, e_j \rangle e_j, e_k \right\rangle = \langle x, e_k \rangle - \langle x, e_k \rangle \langle e_k, e_k \rangle = 0.$$

□

Corollary 10.3.5. *Let I be any index set and $\{e_j\}_{j \in I}$ be an orthonormal set of a Hilbert space X over \mathbf{F} . Then $S = \{e_j : \langle x, e_j \rangle \neq 0\}$ is countable for any $x \in X$.*

Proof. For each $n \in \mathbf{N}$, define $S_n = \{e_j : |\langle x, e_j \rangle|^2 > n^{-1} \|x\|^2\}$. By Lemma 10.3.4 each S_n contains at most $n - 1$ elements. Since $S = \bigcup_{n=1}^{\infty} S_n$, we conclude that S is countable. \square

Remark 10.3.6. *Given an arbitrary orthonormal set $\{e_j\}_{j \in I}$ indexed by a set I , under Lemma 10.3.4 and Corollary 10.3.5, the sums $\sum_{j \in I} |\langle x, e_j \rangle|^2$ and $\sum_{j \in I} \langle x, e_j \rangle e_j$ can be written as the series $\sum_{k=1}^{\infty} |\langle x, e_{j_k} \rangle|^2$ and $\sum_{k=1}^{\infty} \langle x, e_{j_k} \rangle e_{j_k}$ respectively, where we restrict ourselves to the countable number of e_{j_k} for which $\langle x, e_{j_k} \rangle \neq 0$.*

The following result assures that the two series in Remark 10.3.6 are well defined.

Theorem 10.3.7. *Let $\{x_k\}_{k=1}^{\infty}$ be an orthonormal sequence in a Hilbert space X over \mathbf{F} and $\{c_k\}_{k=1}^{\infty} \subseteq \mathbf{F}$. Then $\sum_{k=1}^{\infty} c_k x_k$ is convergent in X if and only if*

$$\left\| \sum_{k=1}^{\infty} c_k x_k \right\|^2 = \sum_{k=1}^{\infty} |c_k|^2 < \infty.$$

Moreover, $\sum_{k=1}^{\infty} c_k x_k$ is independent of the order in which its terms are arranged.

Proof. From the orthonormality of $\{x_k\}_{k=1}^{\infty}$ it follows that for $m, n \in \mathbf{N}$ and $m > n$,

$$\left\| \sum_{k=n}^m c_k x_k \right\|^2 = \sum_{k=n}^m |c_k|^2.$$

This, together with the completeness of X , implies the if and only if part of the theorem.

To complete the proof, let $\sum_{k=1}^{\infty} |c_k|^2 < \infty$ and $y = \sum_{l=1}^{\infty} c_{k_l} x_{k_l}$ be a rearrangement of $x = \sum_{k=1}^{\infty} c_k x_k$. Then

$$\begin{cases} \|x - y\|^2 = \langle x, x \rangle + \langle y, y \rangle - \langle x, y \rangle - \langle y, x \rangle; \\ \langle x, x \rangle = \langle y, y \rangle = \sum_{k=1}^{\infty} |c_k|^2. \end{cases}$$

Note that

$$\begin{cases} s_m = \sum_{k=1}^m c_k x_k \text{ and } t_m = \sum_{l=1}^m c_{k_l} x_{k_l} \Rightarrow \langle x, y \rangle = \lim_{m \rightarrow \infty} \langle s_m, t_m \rangle = \sum_{k=1}^{\infty} |c_k|^2; \\ \langle y, x \rangle = \overline{\langle x, y \rangle} = \langle x, y \rangle. \end{cases}$$

So, it follows that $\|x - y\| = 0$ and hence $x = y$. \square

Here is a statement of Friedrich Wilhelm Bessel about the coefficients of an element in a Hilbert space with respect to an orthonormal set.

Theorem 10.3.8. *Let I be any index set and $\{e_j\}_{j \in I}$ be an orthonormal set of a Hilbert space X over \mathbf{F} . Then there hold the following two statements.*

(i) *Bessel's inequality: $\sum_{j \in I} |\langle x, e_j \rangle|^2 \leq \|x\|^2$ for all $x \in X$.*

(ii) *$\langle x - \sum_{j \in I} \langle x, e_j \rangle e_j, e_k \rangle = 0$ for all $x \in X$ and $k \in I$.*

Proof. (i) This follows from

$$\sum_{j \in I} |\langle x, e_j \rangle|^2 = \lim_{n \rightarrow \infty} \sum_{k=1}^n |\langle x, e_{j_k} \rangle|^2 \leq \|x\|^2.$$

(ii) Using the continuity of the inner product in its left-hand variable that follows from the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} \left\langle x - \sum_{j \in I} \langle x, e_j \rangle e_j, e_k \right\rangle &= \langle x, e_k \rangle - \left\langle \sum_{j \in I} \langle x, e_j \rangle e_j, e_k \right\rangle \\ &= \langle x, e_k \rangle - \left\langle \lim_{n \rightarrow \infty} \sum_{m=1}^n \langle x, e_{j_m} \rangle e_{j_m}, e_k \right\rangle \\ &= \langle x, e_k \rangle - \lim_{n \rightarrow \infty} \left\langle \sum_{m=1}^n \langle x, e_{j_m} \rangle e_{j_m}, e_k \right\rangle \\ &= \langle x, e_k \rangle - \langle x, e_k \rangle = 0. \end{aligned}$$

□

Theorem 10.3.8 (ii) induces the concept of an orthonormal basis for a Hilbert space.

Definition 10.3.9. *Let I be an index set and X be a Hilbert space over \mathbf{F} . Then an orthonormal set $\{e_j\}_{j \in I}$ of X is called an orthonormal basis for X provided that $x = 0$ follows from $\langle x, e_j \rangle = 0$ for all $j \in I$.*

This definition actually illustrates that an orthonormal set is an orthonormal basis when it is impossible to add one more nonzero element to the set while still preserving its orthonormality. An orthonormal basis is not generally a “basis”, i.e., it is not generally possible to write every member of the space as a linear combination of finitely many members of an orthonormal basis. In the infinite-dimensional case the distinction really matters; that is to say, the definition given

above requires only that the span of an orthonormal basis be dense in the normed linear space, not that it equal the entire space. More precisely, we have the following theorem about series by the Marc-Antoine Parseval.

Theorem 10.3.10. *Given an index set I , let $\{e_j\}_{j \in I}$ be an orthonormal set in a Hilbert space X over \mathbf{F} . Then the following statements are equivalent.*

- (i) $\{e_j\}_{j \in I}$ is an orthonormal basis.
- (ii) The closed linear span of $\{e_j\}_{j \in I}$ is X .
- (iii) Parseval's identity: $\|x\|^2 = \sum_{j \in I} |\langle x, e_j \rangle|^2$ for all $x \in X$.

Proof. First, we prove that (i) implies (ii) and (iii). If (i) is true, then Definition 10.3.9 and Theorem 10.3.8 (ii) yield $x = \sum_{j \in I} \langle x, e_j \rangle e_j$ for all $x \in X$. Clearly, (iii) follows from Theorem 10.3.7.

Next, we prove that each of (ii) and (iii) implies (i).

(ii) \Rightarrow (i) Suppose that $y \in X$ satisfies $\langle y, e_j \rangle = 0$ for all $j \in I$. To prove $y = 0$, consider $S = \{x \in X : \langle y, x \rangle = 0\}$. It is easy to see that S is a linear subspace of X . Since $e_j \in S$, it follows that S must contain the linear span of $\{e_j\}_{j \in I}$. On the other hand, S is closed in view of the continuity of the inner product, and so S must contain the closure of the linear span of $\{e_j\}_{j \in I}$. Hence $S = X$ by (ii). In particular, we have $y \in S$ and so $\langle y, y \rangle = 0$, whence $y = 0$ as required.

(iii) \Rightarrow (i) Suppose on the contrary that $\{e_j\}_{j \in I}$ does not form an orthonormal basis of X . Then there exists a nonzero $x \in X$ such that $\langle x, e_j \rangle = 0$ for all $j \in I$. Then

$$0 \neq \|x\|^2 = \sum_{j \in I} |\langle x, e_j \rangle|^2 = 0.$$

This is a contradiction. □

Example 10.3.11.

- (i) The set $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ forms an orthonormal basis of \mathbf{R}^3 .
- (ii) The set $\{e_j : j \in \mathbf{N}\}$ with the j -th entry of $e_j \in \ell_2$ being 1 and others being 0 forms an orthonormal basis of the real sequence space ℓ_2 .
- (iii) The set $\{e^{int}\}_{n=-\infty}^{\infty}$ is an orthonormal basis for $LRS_{id}^2([0, 2\pi], \mathbf{C})$. So, $C([0, 2\pi], \mathbf{C})$ is dense in $LRS_{id}^2([0, 2\pi], \mathbf{C})$. In fact, this basis is fundamental to the study of Fourier series, named in honour of Joseph Fourier.

Theorem 10.3.12.

(i) Every Hilbert space over \mathbf{F} has an orthonormal basis.

(ii) Any orthonormal basis in a separable Hilbert space over \mathbf{F} is countable.

Proof. (i) Let X be a Hilbert space over \mathbf{F} with the inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \|$, and consider the collection E of orthonormal subsets of X . It is clear that E is nonempty and can be partially ordered under inclusion. If F is any totally ordered subcollection of E , the set $U = \cup_{S \in F} S$ is member of E and an upper bound for F . By Zorn's lemma there is a maximal orthonormal set M . Since M is maximal, we conclude from Definition 10.3.9 that M is an orthonormal basis – otherwise, there would be a nonzero $x_0 \in X \setminus M$ such that $\langle x_0, e_j \rangle = 0$ holds for any $e_j \in M$. Accordingly, $\{\|x_0\|^{-1}x_0\} \cup M$ forms an orthonormal set which properly contains M – this contradicts the maximality of M .

(ii) If X is separable, and if $\{e_j\}_{j \in I}$ is an uncountable orthonormal basis indexed by I , then for $j, k \in I$ with $j \neq k$, we have

$$\|e_j - e_k\|^2 = \|e_j\|^2 + \|e_k\|^2 = 2,$$

and so the open balls $B_{2^{-1}}^X(e_j)$ (with centre e_j and radius 2^{-1}) in X are mutually disjoint. If $\{x_j\}_{j=1}^\infty$ is a countable dense subset of X , because $\{e_j\}_{j \in I}$ is uncountable, there is an open ball $B_{2^{-1}}^X(e_{j_0})$ which is disjoint from $\{x_j\}_{j=1}^\infty$. Hence e_{j_0} is not in the closure of $\{x_j\}_{j=1}^\infty$. This contradicts the density of $\{x_j\}_{j=1}^\infty$ in X . We are done. \square

Example 10.3.13. There is a non-separable Hilbert space. As a matter of fact, let X be the linear space of all continuous functions $f : \mathbf{R} \rightarrow \mathbf{C}$ satisfying

$$\|f\|_2 = \left(\lim_{n \rightarrow \infty} (2n)^{-1} \int_{-n}^n |f(x)|^2 dx \right)^{2^{-1}} < \infty.$$

Although $\|f\|_2 = 0$ does not ensure $f = 0$ on \mathbf{R} , if $Y = \{f \in X : \|f\|_2 = 0\}$ then X/Y is a normed linear space over \mathbf{C} . If H is the completion of this quotient space and is equipped with the following inner product

$$\langle f, g \rangle = \lim_{n \rightarrow \infty} (2n)^{-1} \int_{-n}^n f(x) \overline{g(x)} dx,$$

then H is a Hilbert space over \mathbf{C} . Note that $\{e^{it(\cdot)} + Y\}_{t \in \mathbf{R}}$ is an orthonormal set of H but also

$$\{f + Y \in H : \|f(\cdot) - e^{it(\cdot)}\|_2 < 2^{-1}\} \text{ and } \{f + Y \in H : \|f(\cdot) - e^{is(\cdot)}\|_2 < 2^{-1}\}$$

are disjoint for $t \neq s$. So H is not separable.

Corollary 10.3.14. *Any two infinite dimensional separable Hilbert spaces over \mathbf{F} are isometrically isomorphic.*

Proof. Suppose that X and Y are two such spaces equipped with the inner products $\langle \cdot, \cdot \rangle_X$ and $\langle \cdot, \cdot \rangle_Y$. Theorem 10.3.12 tells us that there are sequences $\{x_j\}_{j=1}^\infty$ and $\{y_j\}_{j=1}^\infty$ that form orthonormal bases for X and Y respectively. If $x \in X$ and $y \in Y$, then

$$x = \sum_{j=1}^{\infty} \langle x, x_j \rangle_X x_j \quad \text{and} \quad y = \sum_{j=1}^{\infty} \langle y, y_j \rangle_Y y_j.$$

Define a mapping

$$T : X \rightarrow Y \quad \text{by} \quad T(x) = y \quad \text{whenever} \quad \langle x, x_j \rangle_X = \langle y, y_j \rangle_Y.$$

It is clear that T is linear and one-to-one, and it maps X onto Y since $(\langle x, x_j \rangle_X)_{j=1}^\infty$ and $(\langle y, y_j \rangle_Y)_{j=1}^\infty$ run through all of elements in the \mathbf{F} -valued sequence space $\ell_2(\mathbf{N}, \mathbf{F})$. In addition,

$$\|T(x)\|^2 = \sum_{j=1}^{\infty} |\langle y, y_j \rangle_Y|^2 = \sum_{j=1}^{\infty} |\langle x, x_j \rangle_X|^2 = \|x\|^2,$$

Thus, T is isometrically isomorphic. The proof is complete. \square

10.4 Five Special Bounded Operators

For a Hilbert space X over \mathbf{F} , the elements in $\mathcal{B}(X)$ are of particular interest. To see this, suppose that $T \in \mathcal{B}(X)$. For each fixed $y \in X$, the mapping that sends x to $\langle T(x), y \rangle$ is a continuous linear functional on X . Thus by the Riesz representation theorem there exists a unique $z \in X$ such that

$$\langle T(x), y \rangle = L_z(x) = \langle x, z \rangle.$$

This pairing actually produces a new operator on X – see also Theorem 9.5.6.

Definition 10.4.1. *Let X be a Hilbert space over \mathbf{F} . If $T \in \mathcal{B}(X)$ then its adjoint operator T^* is defined by*

$$\langle T(x), y \rangle = \langle x, T^*(y) \rangle \quad \text{for all } x, y \in X.$$

The three facts (i)-(ii)-(iii) in Theorem 10.4.2 below are the required conditions for $*$ to be an involution.

Theorem 10.4.2. *Let X be a Hilbert space over \mathbf{F} . Then the operation $*$ maps $\mathcal{B}(X)$ to itself, and has the following properties for all $T, S \in \mathcal{B}(X)$ and $\alpha, \beta \in \mathbf{F}$:*

$$(i) (\alpha T + \beta S)^* = \bar{\alpha} T^* + \bar{\beta} S^*;$$

$$(ii) (TS)^* = S^* T^*;$$

$$(iii) T^{**} = T;$$

$$(iv) \|T^*\| = \|T\|;$$

$$(v) \|T^* T\| = \|T\|^2.$$

Proof. First of all, we verify that $*$ maps $\mathcal{B}(X)$ to $\mathcal{B}(X)$. The linearity of T^* follows from the following calculation:

$$\begin{aligned} \langle x, T^*(\alpha y_1 + \beta y_2) \rangle &= \langle T(x), \alpha y_1 + \beta y_2 \rangle \\ &= \bar{\alpha} \langle T(x), y_1 \rangle + \bar{\beta} \langle T(x), y_2 \rangle \\ &= \langle x, \alpha T^*(y_1) + \beta T^*(y_2) \rangle \end{aligned}$$

By the definition of the operator norm, we have

$$\begin{aligned} \|T^*\| &= \sup_{\|y\|=1} \|T^*(y)\| \\ &\leq \sup_{\|x\|=1, \|y\|=1} |\langle x, T^*(y) \rangle| \\ &= \sup_{\|x\|=1, \|y\|=1} |\langle T(x), y \rangle| \\ &\leq \sup_{\|x\|=1} \|T(x)\| \\ &= \|T\|. \end{aligned}$$

This implies that $*$ is bounded on $\mathcal{B}(X)$.

Next, we check those five properties.

(i) For any $x, y \in X$, we have

$$\langle x, (\alpha T + \beta S)^*(y) \rangle = \alpha \langle T(x), y \rangle + \beta \langle S(x), y \rangle = \langle x, (\bar{\alpha} T^* + \bar{\beta} S^*)(y) \rangle.$$

(ii) This follows from

$$\langle x, (TS)^*(y) \rangle = \langle TS(x), y \rangle = \langle S(x), T^*(y) \rangle = \langle x, S^* T^*(y) \rangle.$$

(iii) This follows from

$$\langle x, T^{**}(y) \rangle = \langle T^*(x), y \rangle = \overline{\langle y, T^*(x) \rangle} = \overline{\langle T(y), x \rangle} = \langle x, T(y) \rangle.$$

(iv) It is known that $\|T^*\| \leq \|T\|$ and $T^{**} = T$. So $\|T\| \leq \|T^*\|$. This gives $\|T\| = \|T^*\|$.

(v) By (iv), we get

$$\|T^*T\| \leq \|T^*\| \|T\| = \|T\|^2.$$

For the reverse inequality, we note that

$$\|T(x)\|^2 = \langle T^*T(x), x \rangle \leq \|T^*T(x)\| \|x\| \leq \|T^*T\| \|x\|^2.$$

So, $\|T\|^2 \leq \|T^*T\|$, which completes the proof. \square

Example 10.4.3. If T is the operator on $\ell_2(\mathbf{N}, \mathbf{C})$ defined by

$$T(x_1, x_2, \dots) = (0, x_1, x_2, \dots),$$

then $T^*(x_1, x_2, \dots) = (x_2, x_3, \dots)$. Clearly, $\|T\| = \|T^*\| = 1$.

In what follows, we deal with five special types of bounded linear operators on a given Hilbert space. They are self-adjoint, normal, nonnegative, unitary, and projective operators.

Definition 10.4.4. Let X be a Hilbert space over \mathbf{F} . An operator $T \in \mathcal{B}(X)$ is said to be a self-adjoint or symmetric operator provided that $T = T^*$.

Example 10.4.5.

(i) On a finite-dimensional Hilbert space over \mathbf{F} , a self-adjoint operator is one that is its own adjoint, or, equivalently, one whose matrix is Hermitian, where a Hermitian matrix is one which is equal to its own conjugate transpose.

(ii) Given $n \in \mathbf{N}$, the operator T defined by $T(f)(x) = x^n f(x)$ is a self-adjoint operator on $LRS_{id}^2([0, 1], \mathbf{C})$.

The structure of self-adjoint operators on infinite dimensional Hilbert spaces is complicated somewhat by the following characterization and its immediate remark.

Theorem 10.4.6. Let X be a Hilbert space over \mathbf{F} and $T \in \mathcal{B}(X)$.

(i) If T is self-adjoint, then $\|T\| = \sup_{\|x\|=1} |\langle T(x), x \rangle|$.

(ii) If $F = \mathbf{C}$, then T is self-adjoint when and only when $\langle T(x), x \rangle$ is real for all $x \in X$. In particular, $T = 0$ when and only when $\langle T(x), x \rangle = 0$ for all $x \in X$.

Proof. (i) The Cauchy-Schwartz inequality implies $\|T\| \geq \sup_{\|x\|=1} |\langle T(x), x \rangle|$. For the reversed estimate, we use the triangle inequality and the parallelogram law to obtain that for any $x \in X$ with $\|x\| = 1$,

$$\begin{aligned} 2|\langle T(x), y \rangle + \langle T(y), x \rangle| &= |\langle T(x+y), x+y \rangle - \langle T(x-y), x-y \rangle| \\ &\leq 2 \sup_{\|z\|=1} |\langle T(z), z \rangle| (\|x+y\|^2 + \|x-y\|^2) \\ &= 4 \sup_{\|z\|=1} |\langle T(z), z \rangle| (1 + \|y\|^2). \end{aligned}$$

If $T(x) \neq 0$, taking $y = T(x)\|T(x)\|^{-1}$ in the above we get $\|T(x)\| \leq \sup_{\|z\|=1} |\langle T(z), z \rangle|$. Note that the last inequality is trivially valid for $T(x) = 0$. So, it follows that $\|T\| \leq \sup_{\|x\|=1} |\langle T(x), x \rangle|$, and the desired equality occurs.

(ii) If $T \in \mathcal{B}(X)$ is self-adjoint, then

$$\langle T(x), x \rangle = \langle x, T(x) \rangle = \overline{\langle T(x), x \rangle} \text{ for all } x \in X.$$

Hence $\langle T(x), x \rangle$ is real for all $x \in X$. Conversely, if $f(x) = \langle T(x), x \rangle$, then

$$\begin{cases} f(x+y) = f(x) + f(y) + \langle T(y), x \rangle + \langle T(x), y \rangle; \\ f(x+iy) = f(x) + f(y) + i\langle T(y), x \rangle - i\langle T(x), y \rangle. \end{cases}$$

Since f is real-valued, we conclude from the last two identities that there are $r, s \in \mathbf{R}$ such that

$$\langle T(y), x \rangle + \langle T(x), y \rangle = r \text{ and } \langle T(y), x \rangle - \langle T(x), y \rangle = is.$$

This yields

$$\begin{cases} \langle T(y), x \rangle = 2^{-1}(r+is) \text{ and } \langle T(x), y \rangle = 2^{-1}(r-is); \\ \langle T(y), x \rangle = \overline{\langle T(x), y \rangle} = \overline{\langle x, T^*(y) \rangle} = \langle T^*(y), x \rangle. \end{cases}$$

Consequently, for any $y \in X$, we have

$$\langle (T - T^*)(y), (T - T^*)(y) \rangle = 0, \text{ i.e., } T = T^*.$$

For the second half of (ii), it is enough to prove the sufficiency. Suppose that $\langle T(x), x \rangle = 0$ for all $x \in X$. Then T is self-adjoint, and hence

$$4\langle T(x), y \rangle = \langle T(x+y), x+y \rangle - \langle T(x-y), x-y \rangle = 0$$

holds for all $y = T(x) \in X$. This gives $T(x) = 0$ for all $x \in X$, i.e., $T = 0$. \square

Remark 10.4.7. Unfortunately, Theorem 10.4.6 (i) is not valid for any real Hilbert space. For instance, if $X = \mathbf{R}^2$ and

$$T(x) = T(x_1, x_2) = (-x_2, x_1) \text{ for all } x = (x_1, x_2) \in \mathbf{R}^2$$

then $\langle T(x), x \rangle = 0$, $0 \neq T \in \mathcal{B}(\mathbf{R}^2)$ and T is not self-adjoint since

$$T^*(x) = T^*(x_1, x_2) = (x_2, -x_1) \text{ for all } x = (x_1, x_2) \in \mathbf{R}^2.$$

Definition 10.4.8. Let X be a Hilbert space over \mathbf{C} . An operator $T \in \mathcal{B}(X)$ is said to be a nonnegative operator, denoted $T \geq 0$, provided that $\langle T(x), x \rangle \geq 0$ for all $x \in X$. Moreover, we say that $T_1 \leq T_2$ if T_1 and T_2 are self-adjoint operators in $\mathcal{B}(X)$ and $T_2 - T_1 \geq 0$.

Example 10.4.9.

(i) Any bounded operator T on a given Hilbert space over \mathbf{C} always generates a nonnegative operator T^*T since $\langle T^*T(x), x \rangle = \|T(x)\|^2 \geq 0$.

(ii) The operator defined in Example 10.4.5 (ii) is nonnegative.

(iii) The operator T on \mathbf{C}^2 determined by $T(x_1, x_2) = (x_1 + 2x_2, 2x_1 + x_2)$ is not nonnegative since $T(1, -1) = (-1, 1)$ and the inner product between two vectors $(-1, 1)$ and $(1, -1)$ is -2 .

According to Theorem 10.4.6 (ii) and Example 10.4.9 (ii), any nonnegative operator is self-adjoint, but not conversely. Here is the basic structure of a nonnegative operator.

Theorem 10.4.10. Given a Hilbert space X over \mathbf{C} , let $I \in \mathcal{B}(X)$ be the identity operator.

(i) If T is self-adjoint with $-I \leq T \leq I$ or $0 \leq T \leq 0$, then $\|T\| \leq 1$ or $T = 0$.

(ii) If $T \geq 0$, then $T + I$ is invertible, and the generalized Cauchy-Schwartz inequality holds:

$$|\langle T(x), y \rangle| \leq \sqrt{\langle T(x), x \rangle \langle T(y), y \rangle} \text{ for all } x, y \in X.$$

(iii) If $0 \leq T_1 \leq T_2 \leq \cdots \leq I$, then there exists a $T_\infty \in \mathcal{B}(X)$ such that

$$\lim_{j \rightarrow \infty} \|T_j(x) - T_\infty(x)\| = 0 \text{ for all } x \in X.$$

(iv) $T \geq 0$ when and only when there exists a unique nonnegative operator $S \in \mathcal{B}(X)$, denoted $S = \sqrt{T}$, such that $S^2 = T$.

Proof. (i) If T is self-ad-joint, then $I \pm T \geq 0$ implies $\pm \langle T(x), x \rangle \leq \|x\|^2$ and $\|T\| \leq 1$ by Theorem 10.4.6 (i). Moreover, $\pm T \geq 0$ implies $\langle T(x), x \rangle = 0$ and $\|T\| = 0$ by Theorem 10.4.6 (i).

(ii) Since $T \geq 0$, we conclude

$$(T + I)(x) = 0 \Rightarrow \langle (T + I)(x), x \rangle = 0 \Rightarrow 0 \leq \langle T(x), x \rangle = -\|x\|^2 \Rightarrow x = 0,$$

and so the injectivity of $T + I$ follows. The surjectivity will follow if we can prove that $M = (T + I)(X)$, the image of $T + I$, is both dense and closed. The argument used for injectivity applies to $x \in M^\perp$ to infer that $\langle (T + I)x, x \rangle = 0$ for all $x \in M^\perp$ and so $M^\perp = \{0\}$. On the other hand, for any $x \in X$, we have

$$\|(T + I)(x)\|^2 = \|T(x)\|^2 + 2\langle T(x), x \rangle + \|x\|^2$$

and hence $\|x\| \leq \|(T + I)(x)\|$. Consequently, if $\{(T + I)(x_j)\}_{j=1}^\infty$ is a Cauchy sequence in M , then $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence in X and hence it is convergent. This shows that M is complete and thereby closed in X . From Theorem 10.2.5 it turns out that $X = M \oplus M^\perp = M$; that is, $T + I$ is surjective.

Regarding the generalized Cauchy-Schwartz inequality for any $x, y \in X$, for any $t \in \mathbf{R}$ let $\lambda = t\zeta \in \mathbf{F}$ obey $\zeta \langle y, T(x) \rangle = |\langle y, T(x) \rangle|$. Then we get

$$0 \leq \langle x + \lambda y, T(x + \lambda y) \rangle = \langle x, T(x) \rangle + 2t|\langle y, T(x) \rangle| + t^2 \langle y, T(y) \rangle,$$

whence deriving

$$|\langle T(x), y \rangle|^2 - \langle x, T(x) \rangle \langle y, T(y) \rangle \leq 0.$$

(iii) For $k > j$ and $x \in X$, we use the generalized Cauchy-Schwartz inequality in (ii) to get

$$\begin{aligned} \|T_k(x) - T_j(x)\|^2 &= \langle (T_k - T_j)(x), (T_k - T_j)(x) \rangle \\ &\leq \sqrt{\langle (T_k - T_j)(x), x \rangle \langle (T_k - T_j)^2(x), (T_k - T_j)(x) \rangle} \\ &\leq \sqrt{\langle (T_k - T_j)(x), x \rangle} \|x\| \\ &\leq \sqrt{\langle T_k(x), x \rangle - \langle T_j(x), x \rangle} \|x\| \end{aligned}$$

since $0 \leq T_k - T_j \leq I$ implies $\|T_k - T_j\| \leq 1$ and $\|(T_k - T_j)^2\| \leq 1$. Note that $\{\langle T_j(x), x \rangle\}_{j=1}^\infty$ is convergent since it is increasing and bounded. So $\{T_j(x)\}_{j=1}^\infty$ is a Cauchy sequence in the Hilbert space X and consequently, this sequence is convergent to $T_\infty(x) \in X$. According to Theorem 9.2.4 (i), $T_\infty \in \mathcal{B}(X)$.

(iv) The sufficiency is trivial. For the necessity, suppose $T \geq 0$. According to Theorem 10.4.6 (i) we have $\|T\| < \infty$ and equivalently $T \leq \|T\|I$. Therefore, without loss of generality we may assume $T \leq I$. To find a nonnegative operator S such that $S^2 = T$, let $U = I - T$, $T_1 = 0$ and $T_{j+1} = 2^{-1}(U + T_j^2)$ for $j \in \mathbf{N}$. By induction, we find three basic facts as follows:

- (a) For $j \in \mathbf{N}$, $0 \leq T_j$ is a polynomial of U with nonnegative coefficients;
- (b) For $j \in \mathbf{N}$, $T_j \leq I$. In fact, $T_j \leq I$ implies

$$\langle T_j^2(x), x \rangle = \|T_j(x)\|^2 \leq \|x\|^2$$

and so $T_j^2 \leq I$ which, along with $U \leq I$, yields $T_{j+1} \leq I$;

(c) For $j \in \mathbf{N}$, $T_{j+2} - T_{j+1} \geq 0$. This is because the easy commutativity $T_{j+1}T_j = T_jT_{j+1}$ deduces

$$T_{j+2} - T_{j+1} = 2^{-1}(T_{j+1}^2 - T_j^2) = 2^{-1}(T_{j+1} + T_j)(T_{j+1} - T_j),$$

where $T_{j+1} + T_j$ is a polynomial of U with nonnegative coefficients, and so is $T_{j+1} - T_j$ by $T_2 - T_1 = 2^{-1}U$ and by induction.

Now, (iii) is applied to produce a $T_\infty \in \mathcal{B}(X)$ with $\lim_{j \rightarrow \infty} T_j(x) = T_\infty(x)$ for all $x \in X$. Then $T_\infty = 2^{-1}(U + T_\infty^2)$ and hence desired square root operator is $S = I - T_\infty$, i.e., $(I - T_\infty)^2 = T$.

It remains to check the uniqueness. Suppose that $S_1 \geq 0$ satisfies $S_1^2 = T$. Then $S_1T = S_1^3 = TS_1$ and hence $S_1T^n = T^nS_1$ for any $n \in \mathbf{N}$. Now the previous-established limit process implies $SS_1 = S_1S$. Consequently, if $y = (S - S_1)(x)$ for any $x \in X$ then $0 = \langle (S + S_1)(y), y \rangle = \langle S(y), y \rangle + \langle S_1(y), y \rangle$, and hence $S(y) = 0 = S_1(y)$ (here $S = \sqrt{S^2}$ and $S_1 = \sqrt{S_1^2}$) which ensures

$$(S^2 - SS_1)(x) = 0 = (S_1S - S_1^2)(x).$$

Accordingly,

$$\|(S - S_1)(x)\|^2 = \langle (S - S_1)^2(x), x \rangle = 0,$$

and so $S = S_1$. □

Definition 10.4.11. Given a Hilbert space X over \mathbf{F} let $T \in \mathcal{B}(X)$. Then T is called unitary provided that T is an isometric isomorphism.

Example 10.4.12. For $f \in LRS_{id}^2(\mathbf{R}, \mathbf{C})$ let

$$\mathcal{F}(f)(x) = \sqrt{2\pi}^{-1} \int_{\mathbf{R}} f(t) \exp(itx) dm_{id}(t) \text{ for all } x \in \mathbf{R},$$

then \mathcal{F} is a unitary operator on $LRS_{id}^2(\mathbf{R}, \mathbf{C})$. This operator is called the Fourier transform and its adjoint operator is determined by

$$\mathcal{F}^*(f)(x) = \sqrt{2\pi}^{-1} \int_{\mathbf{R}} f(t) \exp(-itx) dm_{id}(t) \text{ for all } x \in \mathbf{R}.$$

We can characterize unitary operators in terms of adjoints and inner products.

Theorem 10.4.13. Given a Hilbert space X over \mathbf{C} let $T \in \mathcal{B}(X)$. If I is the identity operator in $\mathcal{B}(X)$, then the following statements are equivalent:

- (i) T is unitary;
- (ii) $\langle T(x), T(y) \rangle = \langle x, y \rangle$ for all $x, y \in X$, and T is surjective;
- (iii) $TT^* = T^*T = I$.

Proof. (i) \Rightarrow (ii) If T is unitary, then $\|T(x)\| = \|x\|$ and hence

$$\begin{aligned} \langle T(x), T(y) \rangle &= 2^{-2} \sum_{n=1}^4 i^n \|T(x) + i^n T(y)\|^2 \\ &= 2^{-2} \sum_{n=1}^4 i^n \|T(x + i^n y)\|^2 \\ &= 2^{-2} \sum_{n=1}^4 i^n \|x + i^n y\|^2 \\ &= \langle x, y \rangle. \end{aligned}$$

The surjectivity of T follows the definition of a unitary operator.

(ii) \Rightarrow (iii) Clearly, (ii) implies that T is injective. Moreover

$$\langle T^*T(x), x \rangle = \langle T(x), T(x) \rangle = \langle x, x \rangle.$$

This, along with Theorem 10.4.6 (ii), yields

$$\langle (T^*T - I)(x), x \rangle = 0 \Rightarrow T^*T = I \Rightarrow TT^* = I.$$

(iii) \Rightarrow (i) If (iii) is true, then T is surjective and

$$\|T(x)\|^2 = \langle T(x), T(x) \rangle = \langle x, T^*T(x) \rangle = \langle x, x \rangle = \|x\|^2.$$

This implies that T is injective, and thereby T is unitary. \square

Theorem 10.4.13 leads to the following consideration.

Definition 10.4.14. Let X be a Hilbert space over \mathbf{C} . An operator $T \in \mathcal{B}(X)$ is said to be a normal operator if $TT^* = T^*T$.

Interestingly, the normal operators correspond to the complex numbers.

Theorem 10.4.15. Let X be a Hilbert space over \mathbf{C} and $T \in \mathcal{B}(X)$. Then the following are equivalent.

- (i) T is normal.
- (ii) $T = T_1 + iT_2$ where T_1 and T_2 are self-adjoint and $T_1T_2 = T_2T_1$.
- (iii) $\|T(x)\| = \|T^*(x)\|$ for all $x \in X$.

Proof. (i) \Rightarrow (ii) Put

$$T_1 = 2^{-1}(T + T^*) \quad \text{and} \quad T_2 = (2i)^{-1}(T - T^*).$$

It is easy to see that $T = T_1 + iT_2$ holds with T_1 and T_2 being self-adjoint and commutative.

(ii) \Rightarrow (iii) Using the given decomposition of T plus $T_1T_2 = T_2T_1$, we obtain

$$\begin{aligned} \|T(x)\|^2 &= \langle x, T^*T(x) \rangle \\ &= \langle x, (T_1 - iT_2)(T_1 + iT_2)(x) \rangle \\ &= \langle x, (T_1T_1 + T_2T_2)(x) \rangle \\ &= \langle x, (T_1 + iT_2)(T_1 - iT_2)(x) \rangle \\ &= \langle x, TT^*(x) \rangle \\ &= \|T^*(x)\|^2. \end{aligned}$$

(iii) \Rightarrow (i). For any $x \in X$, we have

$$\langle (TT^* - T^*T)(x), x \rangle = \langle TT^*(x), x \rangle - \langle T^*T(x), x \rangle = \|T^*(x)\|^2 - \|T(x)\|^2 = 0.$$

So, Theorem 10.4.6 is employed to derive that $TT^* - T^*T = 0$, as desired. \square

Example 10.4.16. Let $T = 2il$ on a given Hilbert space X over \mathbf{C} , where l is the identity operator in $\mathcal{B}(X)$. Then $T^* = -2il$ and $TT^* = T^*T = 4l$. Thus, T is normal operator but not unitary nor self-adjoint.

The fifth special operator that we take account of is the projection operator.

Definition 10.4.17. Let X be a Hilbert space over \mathbf{F} . Then an operator $T \in \mathcal{B}(X)$ is called a projection operator provided $T^2 = T$.

Example 10.4.18. Let M be a closed subspace of a Hilbert space X over \mathbf{F} . Then $X = M \oplus M^\perp$; that is, any $x \in X$ there are unique $y \in M$ and $z \in M^\perp$ such that $x = y + z$. If $P_M(x) = y$ then P_M is a projection. In fact,

$$P_M^2(x) = P_M(y) = y = P_M(x) \Rightarrow P_M^2 = P_M.$$

The fact that $P_M \in \mathcal{B}(X)$ follows from

$$\|x\|^2 = \|y + z\|^2 = \|y\|^2 + \|z\|^2 \geq \|y\|^2 = \|P_M(x)\|^2 \Rightarrow \|P_M\| \leq 1.$$

Traditionally, P_M is called the orthogonal projection of X onto M . Moreover, if $M \neq \{0\}$ then $P_M \neq 0$ and hence for any $x \in M \setminus \{0\}$ we have $x = x + 0$ and $P_M(x) = x$, giving $\|P_M\| = 1$.

The following theorem shows that the above orthogonal projections exist as a very important subclass of the projection operators.

Theorem 10.4.19. Let X be a Hilbert space over \mathbf{C} . If $T \in \mathcal{B}(X)$ is a projection, then the following statements are equivalent.

- (i) T is nonnegative.
- (ii) T is self-adjoint.
- (iii) T is normal.
- (iv) T is the orthogonal projection on $T(X)$.

Proof. Since (i) \Rightarrow (ii) \Rightarrow (iii) are straightforward, we only verify (iii) \Rightarrow (iv) \Rightarrow (i). (iii) \Rightarrow (iv) Assume (iii) holds. To reach (iv), let $M = T(X)$.

We first prove that M is closed. Given $\{y_j\}_{j=1}^\infty$ in M with $y_j \rightarrow y$, we have $y_j = T(x_j)$, $x_j \in X$. Since T is a projection, we can conclude from $y_j \rightarrow y$ that

$$y_j = T(x_j) = T^2(x_j) = T(y_j) \rightarrow T(y) \Rightarrow y = T(y) \in M$$

and hence M is closed.

Next, given $x \in X$ write $x = y + z$, $y \in M$ and $z \in M^\perp$. We must verify $T(x) = y$. Because $T(x) = T(y) + T(z)$, it suffices to prove that $T(y) = y$ and $T(z) = 0$. Since $y \in M$, there is $w \in X$ such that $y = T(w)$ and so $T(y) = T^2(w) = T(w) = y$. By the definition of M we have $T(z) \in M$, and if we can also prove that $T(z) \in M^\perp$, then we can conclude that $T(z) = 0$. To do so, given any $u \in M$, then there is $v \in X$ such that $u = T(v)$ and

$$\langle T(z), u \rangle = \langle T(z), T(v) \rangle = \langle z, T^*T(v) \rangle = \langle z, TT^*(v) \rangle = 0$$

due to $z \in M^\perp$ and $TT^*(v) \in M$. In other words, $T(z) \in M^\perp$.

(iv) \Rightarrow (i) For any $x \in X$ let

$$x = y + z, \quad y \in M = T(X) \quad \text{and} \quad z \in M^\perp.$$

Then

$$\langle T(x), x \rangle = \langle y, y + z \rangle = \langle y, y \rangle + \langle y, z \rangle = \langle y, y \rangle \geq 0.$$

Hence T is nonnegative and the proof is complete. \square

Remark 10.4.20. Here it is worth remarking that there exists a natural one-to-one correspondence between projection operators T and direct sum decompositions $X = M \oplus N$ where M and N are closed. As a matter of fact, if T is a projection, then $X = M \oplus N$ where $M = T(X)$ and $N = (I - T)(X)$. Conversely, if $X = M \oplus N$, then we define T as was done for the orthogonal projection but using N instead of M^\perp . The orthogonal projections are exactly those that arise when $N = M^\perp$, and these are generally the projections of interest in Hilbert space theory.

10.5 Compact Operators via Spectrum

This section is devoted to a spectrum-based study of the structure of a compact operator acting on a Hilbert space.

Definition 10.5.1. Let $(X, \|\cdot\|_X)$ be a Banach space over \mathbf{F} . Then a linear operator T is said to be of finite rank provided that it is of the following form:

$$T(x) = \sum_{k=1}^n L_k(x)x_k \quad \text{for } x, x_k \in X \quad \text{and} \quad L_k \in X^*.$$

Clearly, an operator of finite rank is compact, for it takes a bounded sequence into a bounded sequence in a finite dimensional Banach space. Hence, the image of the sequence has a convergent subsequence. By Theorem 9.5.4 we know that in a Banach space the limit in norm of operators of finite rank is compact. So, it is very natural to wonder whether or not the converse is true; that is, for each compact operator, can we find a sequence of operators of finite rank that approximate this compact operator in norm? Although the answer is no in general for Banach spaces, we will show that the answer is yes for all complex Hilbert spaces.

Definition 10.5.2. Let X be a Banach space over \mathbf{C} and $T \in \mathcal{B}(X)$.

(i) The null space of T , denoted $N(T)$, is $\{x \in X : T(x) = 0\}$. The range of T , denoted $R(T)$, is $\{y \in X : y = T(x) \text{ for some } x \in X\}$.

(ii) A complex number λ is called an eigenvalue of T if there exists a nonzero $x \in X$ such that $T(x) = \lambda x$. In this case, x is called an eigenvector associated with λ .

(iii) The spectrum of T , denoted $\sigma(T)$, is the set of all complex numbers λ such that $T - \lambda I$ does not have a bounded inverse where I is the identity operator in $\mathcal{B}(X)$. The resolvent of T , written $\rho(T)$, is $\mathbf{C} \setminus \sigma(T)$. The spectral radius of T , denoted $r_\sigma(T)$, is $\sup\{|\lambda| : \lambda \in \sigma(T)\}$. In addition, $\sigma(T)$ can be split into the following three classes:

(a) The point spectrum

$$\sigma_p(T) = \{\lambda \in \sigma(T) : N(T - \lambda I) \neq \{0\}\}.$$

(b) The continuous spectrum

$$\sigma_c(T) = \{\lambda \in \sigma(T) : N(T - \lambda I) = \{0\}, \overline{R(T - \lambda I)} = X\}.$$

(c) The residual spectrum

$$\sigma_r(T) = \{\lambda \in \sigma(T) : N(T - \lambda I) = \{0\}, \overline{R(T - \lambda I)} \neq X\}.$$

Example 10.5.3.

(i) For the closed interval $[0, 1]$ of \mathbf{R} , consider $LRS_{id}^2([0, 1], \mathbf{C})$ and the differential operator $T = -d^2/dx^2$ defined on the linear subspace of all complex-valued functions f having derivatives of any order on $[0, 1]$ and satisfying the boundary

conditions: $f(0) = f(1) = 0$. Then a computation via integration-by-parts indicates that T is self-adjoint; its eigenfunctions are $f_j(x) = \sin(j\pi x)$, $j \in \mathbf{N}$ with the real eigenvalues $j^2\pi^2$; the well-known orthogonality of the sine functions follows as a consequence of the property of being self-adjoint.

(ii) $\sigma(T)$ is compact. In fact, according to Theorem 9.1.10 (i), $|\lambda| > \|T\|$ implies

$$\lambda(T - \lambda I)^{-1} = (\lambda^{-1}T - I)^{-1} \in \mathcal{B}(X)$$

and then $\lambda \notin \sigma(T)$. By contrapositive statement, if $\lambda \in \sigma(T)$ then $|\lambda| \leq \|T\|$. Thus, $r_\sigma(T) \leq \|T\|$. This implies that $\sigma(T)$ is bounded. To see that $\sigma(T)$ is closed, let $\Phi: \mathbf{C} \rightarrow \mathcal{B}(X)$ be $\Phi(\lambda) = T - \lambda I$. This map is continuous on \mathbf{C} since $\|\Phi(\lambda_1) - \Phi(\lambda_2)\|_X = \|I\||\lambda_1 - \lambda_2|$. Thus, $\sigma(T)$ is closed as the set of all invertible elements of $\mathcal{B}(X)$ is open. Of course, $\rho(T) = \mathbf{C} \setminus \sigma(T)$ is open.

(iii) Let T be defined by $T((x_1, x_2, \dots)) = (x_2, x_3, \dots)$ on $\ell_2(\mathbf{N}, \mathbf{C})$. It is easy to see $\sigma(T) \subseteq \{\lambda \in \mathbf{C} : |\lambda| \leq 1\}$. Since $|\lambda| < 1$ implies $x = (1, \lambda, \lambda^2, \dots) \in \ell_2(\mathbf{N}, \mathbf{C})$ and $T(x) = \lambda x$, and consequently $\{\lambda \in \mathbf{C} : |\lambda| < 1\} \subseteq \sigma_p(T)$. Note that $\sigma(T)$ is closed. So $\{\lambda \in \mathbf{C} : |\lambda| = 1\} \subseteq \sigma(T)$. Also, $\lambda \in \mathbf{C}$ and $|\lambda| = 1$ and $T(x) = \lambda x$ imply $x = (x_1, x_2, \dots) = x_1(1, \lambda, \lambda^2, \dots) \notin \ell_2(\mathbf{N}, \mathbf{C})$ unless $x_1 = 0$. This means $\{\lambda \in \mathbf{C} : |\lambda| = 1\} \subseteq \sigma_c(T)$. Accordingly, $\sigma_r(T) = \emptyset$.

The following result, which will be used later, gives much more information on the spectrum of a self-adjoint operator.

Theorem 10.5.4. *Let X be a Hilbert space over \mathbf{C} and $T \in \mathcal{B}(X)$ be self-adjoint.*

(i) *If λ is an eigenvalue of T , then λ is real.*

(ii) *If x_1 and x_2 are eigenvectors associated with different eigenvalues λ_1 and λ_2 of T , then $\langle x_1, x_2 \rangle = 0$.*

(iii) *If λ is not an eigenvalue of T , then $R(T - \lambda I)$ is dense in X .*

(iv) *$\lambda \in \rho(T)$ if and only if $\|(T - \lambda I)(x)\| \geq \kappa\|x\|$ for all $x \in X$ and some $\kappa > 0$.*

(v) *If $m_T = \inf_{\|x\|=1} \langle T(x), x \rangle$ and $M_T = \sup_{\|x\|=1} \langle T(x), x \rangle$, then $M_T = \|T\|$ and $\sigma(T)$ is a nonempty subset of the closed interval $[m_T, M_T]$. In particular, $m_T, M_T \in \sigma(T)$.*

(vi) $r_\sigma(T) = \|T\|$.

Proof. (i) This follows from the fact that $T(x) = \lambda x$ for some $x \neq 0$ implies

$$\lambda\|x\|^2 = \langle T(x), x \rangle \in \mathbf{R}$$

due to Theorem 10.4.6.

(ii) In this case, by (i) we have

$$\lambda_1 \langle x_1, x_2 \rangle = \langle T(x_1), x_2 \rangle = \langle x_1, T(x_2) \rangle = \lambda_2 \langle x_1, x_2 \rangle$$

which yields $\langle x_1, x_2 \rangle = 0$.

(iii) Assume that λ is not an eigenvalue of T . If $z \in R(T - \lambda I)^\perp$ then

$$0 = \langle z, T(x) - \lambda x \rangle = \langle T(z), x \rangle - \langle \bar{\lambda} z, x \rangle = \langle (T - \bar{\lambda} I)z, x \rangle$$

for all $x \in X$. If $z \neq 0$, then $\bar{\lambda}$ is an eigenvalue of T , and hence $\bar{\lambda}$ is real. Since $T = T^*$, we must then have $\bar{\lambda} = \lambda$, which contradicts the assumption that λ is not an eigenvalue of T . Thus $z = 0$ and

$$\begin{aligned} X &= \overline{R(T - \lambda I)} \oplus \overline{(R(T - \lambda I))^\perp} \\ &\subseteq \overline{R(T - \lambda I)} \oplus R(T - \lambda I)^\perp \\ &= \overline{R(T - \lambda I)} \\ &\subseteq X. \end{aligned}$$

(iv) If $\lambda \in \rho(T)$ then $(T - \lambda I)^{-1} \in \mathcal{B}(X)$ and we may set $\kappa = \|(T - \lambda I)^{-1}\|^{-1}$. Conversely, if $\|(T - \lambda I)(x)\| \geq \kappa \|x\|$ for all $x \in X$ and some $\kappa > 0$, then $R(T - \lambda I)$ is closed. In fact, suppose $(T - \lambda I)(x_k) \rightarrow y$. Then $\{(T - \lambda I)(x_k)\}_{k=1}^\infty$ is a Cauchy sequence and hence $\{x_k\}_{k=1}^\infty$ is a Cauchy sequence too. Since X is Hilbert space, there exists a $z \in X$ such that $x_k \rightarrow z$. By continuity we have $(T - \lambda I)(x_k) \rightarrow T(z)$ and therefore $y = T(z)$. Also, by (iii) above, we see that $R(T - \lambda I)$ is dense in X and so $R(T - \lambda I) = X$. The given inequality insures that $N(T - \lambda I) = \{0\}$. Hence $T - \lambda I$ is surjective and injective. Accordingly, $(T - \lambda I)^{-1} \in \mathcal{B}(X)$, namely, $\lambda \in \rho(T)$.

(v) Clearly, $M_T \leq \|T\|$. To prove $\|T\| \leq M_T$, assume $\|z\| = 1$ and $T(z) \neq 0$. If

$$\begin{cases} b = \|T(z)\|^{\frac{1}{2}}; \\ x = bz + b^{-1}T(z); \\ y = bz - b^{-1}T(z), \end{cases}$$

then by the parallelogram law,

$$\|x\|^2 + \|y\|^2 = 2^{-1}(\|x+y\|^2 + \|x-y\|^2) = 2b^2 + 2b^{-2}\|T(z)\|^2 = 4\|T(z)\|.$$

Note that $|\langle Tu, u \rangle| \leq M_T \|u\|^2$ for any $u \in X$. So, the last equation and a routine calculation give

$$4\|T(z)\|^2 = \langle T(x), x \rangle - \langle x, T(x) \rangle \leq M_T (\|x\|^2 + \|y\|^2) = 4M_T \|T(z)\|,$$

and consequently, $\|T(z)\| \leq M_T$; that is, $\|T\| \leq M_T$.

We next prove that $\sigma(T)$ comprises only real numbers. In so doing, let $\lambda = \alpha + i\beta$ where $\beta \neq 0$. Then for any $x \in X$ we have

$$\|(T - \lambda I)(x)\|^2 = \|T(x) - \alpha x\|^2 + \beta^2 \|x\|^2 \geq \beta^2 \|x\|^2,$$

and so $\lambda \in \rho(T)$ by (iv). In other words, if $\lambda \in \sigma(T)$ then $\beta = 0$ and hence $\lambda \in \mathbf{R}$.

After that, if $\lambda \in (-\infty, m_T)$ then for any $x \in X$,

$$\|T(x) - \lambda x\| \|x\| \geq \langle T(x) - \lambda x, x \rangle = \langle T(x), x \rangle - \lambda \|x\|^2 \geq (m_T - \lambda) \|x\|^2$$

and hence $\lambda \in \rho(T)$ by (iv). Similarly, if $\lambda \in (M_T, \infty)$ then $\lambda \in \rho(T)$. Thus $\sigma(T) \subseteq [m_T, M_T]$.

Furthermore, in order to verify $M_T \in \sigma(T)$, we may assume $M_T \geq m_T \geq 0$ and $M_T = \|T\|$ by replacing T by $T - m_T I$. By the definition of M_T there exists a sequence $\{x_k\}_{k=1}^\infty$ with $\|x_k\| = 1$ such that $\lim_{k \rightarrow \infty} \langle T(x_k), x_k \rangle = M_T$. Accordingly,

$$\|T(x_k) - M_T x_k\|^2 \leq M_T^2 - 2M_T \langle T(x_k), x_k \rangle + M_T^2 \rightarrow 0 \text{ as } k \rightarrow \infty.$$

This, along with (iv), implies $M_T \in \sigma(T)$.

Finally, in order to check $m_T \in \sigma(T)$, we consider $S = -T$ and obtain $-m_T = M_S \in \sigma(S)$ by the previous argument. Thus

$$T - m_T I = -(S - M_S I)$$

does not have a bounded inverse, and hence $m_T \in \sigma(T)$.

(vi) By (v), we have $m_T, M_T \in \sigma(T)$ and so

$$r_\sigma(T) \geq \max\{|m_T|, |M_T|\} = \|T\|.$$

This and the general estimate $r_\sigma(T) \leq \|T\|$ yields the desired formula. \square

When working with a compact self-adjoint operator, we find that the structure of its spectrum is particularly simple.

Theorem 10.5.5. *Let X be a Hilbert space over \mathbf{C} , and let $T \in \mathcal{B}(X)$ be compact and self-adjoint.*

- (i) *If $\lambda \in \sigma(T)$ is nonzero, then λ is an eigenvalue of T .*
(ii) *If $\{e_\alpha : \alpha \in A\}$ is an orthonormal set of eigenvectors corresponding to eigenvalues λ_α with $|\lambda_\alpha| > c > 0$, then A is a finite set.*

Proof. (i) By Theorem 10.5.4 (iv), we see that if $\lambda \in \sigma(T)$ is not equal to 0 then there is a sequence $\{x_k\}_{k=1}^\infty$ in X obeying $\|x_k\| = 1$ and $T(x_k) - \lambda x_k \rightarrow 0$. Since T is compact, there exists a subsequence $\{x_{k_n}\}_{n=1}^\infty$ such that $T(x_{k_n}) \rightarrow y \in X$ as $n \rightarrow \infty$. Consequently,

$$x_{k_n} = \lambda^{-1} (T(x_{k_n}) - (T(x_{k_n}) - \lambda x_{k_n})) \rightarrow \lambda^{-1} y \text{ as } n \rightarrow \infty,$$

and therefore $\lambda y = \lim_{n \rightarrow \infty} \lambda T(x_{k_n}) = T(y)$ owing to the continuity of T ; that is, λ is an eigenvalue of T and y is a corresponding eigenvector. Moreover, $y \neq 0$ since $\|x_k\| = 1$ and $\lambda \neq 0$.

(ii) Assume A is infinite. Then there exists an infinite sequence of orthonormal eigenvectors $\{y_k\}_{k=1}^\infty$ corresponding to eigenvalues $\{\lambda_k\}_{k=1}^\infty$ with $|\lambda_k| > c > 0$ for all $k \in \mathbf{N}$. Note that the orthonormality of $\{y_k\}_{k=1}^\infty$ gives

$$\|T(y_k) - T(y_l)\|^2 = \|\lambda_k y_k\|^2 + \|\lambda_l y_l\|^2 = |\lambda_k|^2 + |\lambda_l|^2 > 2c^2.$$

So, the compactness of T implies $0 \geq c$, a contradiction. \square

To reach the main result of this section, we introduce the following terminology.

Definition 10.5.6. *Let X be a Hilbert space over \mathbf{C} . If λ is an eigenvalue of $T \in \mathcal{B}(X)$, then $N(T - \lambda I)$ is called the eigenspace associated with λ .*

Remark 10.5.7.

(i) *It is clear that $N(T - \lambda I)$ is invariant under T ; that is, if $x \in N(T - \lambda I)$ then $T(x) \in N(T - \lambda I)$ since $T(T(x)) = T(\lambda x) = \lambda T(x)$.*

(ii) *If T is compact and self-adjoint, then the eigenspace corresponding to each nonzero eigenvalue must be finite dimensional, but also, there can be at most finitely many eigenvalues λ obeying $|\lambda| > c$ for a given constant $c > 0$. The former clearly follows from Theorem 10.5.5 (ii). To see the latter, we assume that there were an infinite sequence of different eigenvalues $\{\lambda_k\}_{k=1}^\infty$ obeying*

$|\lambda_k| > c > 0$. If $\{y_k\}_{k=1}^{\infty}$ is a corresponding sequence of norm one eigenvectors, then by Theorem 10.5.4 (ii) we see that $\{y_k\}_{k=1}^{\infty}$ are orthogonal and hence orthonormal. But, $\{y_k\}_{k=1}^{\infty}$ cannot be infinite by Theorem 10.5.5 (ii). Thus, there can be only finitely many eigenvalues obeying $|\lambda_k| > c > 0$. Consequently, the set of nonzero eigenvalues of the operator is at most countable. Moreover, if the operator has infinitely many eigenvalues, then they must converge to zero since at most finitely many eigenvalues can satisfy an inequality of the form $|\lambda| > c > 0$.

Below is the spectral representation of a compact and self-adjoint operator on a complex Hilbert space.

Theorem 10.5.8. *Let X be a Hilbert space over \mathbb{C} , and let $T \in \mathcal{B}(X)$ be compact and self-adjoint. If $\{\lambda_j\}_{j=1}^{\infty}$ is an enumeration of the distinct nonzero eigenvalues of T , then for each $x \in X$ one has $T(x) = \sum_{k=1}^{\infty} \lambda_k P_k(x)$, where each P_k is the orthogonal projection of X onto the eigenspace $N(T - \lambda_k I)$.*

Proof. From the definition of each P_k and Theorem 10.2.2 it follows that for $x \in X$,

$$\|P_k(x) - x\| = \inf\{\|y - x\| : y \in N(T - \lambda_k I)\}.$$

If $\{y_{k,1}, \dots, y_{k,n_k}\}$ is an orthonormal basis for $N(T - \lambda_k I)$, then by Remark 10.5.7 (ii), and Theorems 10.3.8 and 10.3.10, $P_k(x) = \sum_{j=1}^{n_k} \langle x, y_{k,j} \rangle y_{k,j}$. Because eigenvectors corresponding to different eigenvalues are orthogonal, the set $S = \cup_{k=1}^{\infty} \{y_{k,1}, \dots, y_{k,n_k}\}$ is an orthonormal set. The vector

$$y = \sum_{k=1}^{\infty} P_k(x) = \sum_{k=1}^{\infty} \sum_{j=1}^{n_k} \langle x, y_{k,n_k} \rangle y_{k,n_k}$$

is then well-defined since

$$\sum_{k=1}^{\infty} \sum_{j=1}^{n_k} |\langle x, y_{k,n_k} \rangle|^2 \leq \|x\|^2$$

which follows from Bessel's inequality. Since $P_k(x) \in N(T - \lambda_k I)$, we have $T(y) = \sum_{k=1}^{\infty} \lambda_k P_k(x)$. Writing $x = y + (x - y)$, we find that the proof will be complete upon verifying $x - y \in N(T)$. To this end, let M be the smallest closed linear subspace of X containing S . Then S is an orthonormal basis for M and hence $y = P_M(x)$, the orthogonal projection onto M of x . From this it turns out that

$$x = (I - P_M)(x) + P_M(x) = P_{M^\perp}(x) + P_M(x).$$

Since M is invariant under T , $x \in M^\perp$ implies $\langle T(x), z \rangle = \langle x, T(z) \rangle = 0$ for $z \in M$, namely, M^\perp is invariant under T . Therefore, if U stands for the restriction of T to M^\perp , then $U \in \mathcal{B}(M^\perp)$ and it is obviously a compact self-adjoint operator on the Hilbert space M^\perp . Now, if $U \neq 0$, then it has a nonzero eigenvalue which would also be a nonzero eigenvalue of T . In other words, there is some nonzero $x \in M^\perp$ with $x \in N(T - \lambda I) \subseteq M$ for some nonzero $\lambda \in \sigma(T)$. Noticing $M \cap M^\perp = \{0\}$, we get $x = 0$, a contradiction. Therefore, U must be a zero operator and consequently,

$$T(x - y) = UP_{M^\perp}(x) = 0.$$

This completes the proof of the theorem. \square

Remark 10.5.9. *The argument for Theorem 10.5.8 produces a linear subspace M^\perp of X comprising elements mapped to zero by T . So, we can choose an orthonormal basis for M^\perp , each element of which is an eigenvector of T with eigenvalue 0. If these new eigenvectors are included in the sequence of eigenvectors constructed in Theorem 10.5.8, we obtain an orthonormal basis $\{e_j\}$ for X . Note that the new eigenvalues are zero, their inclusion does not affect the convergence of the sequence $\{\lambda_k\}_{k=1}^\infty$. Consequently, if $x = \sum_{j=1}^\infty c_j e_j \in X$ and $\{\lambda_j\}$ are all eigenvalues corresponding to $\{e_j\}$, then*

$$T(x) = \sum_{j=1}^\infty c_j \lambda_j e_j = \sum_{k=1}^\infty c_k \lambda_k P_k(x).$$

In some ways, Theorem 10.5.8 tells us that the self-adjoint compact operators in a Hilbert space are similar to the symmetric matrices in \mathbf{R}^n . Moreover, as an interesting consequence of Theorem 10.5.8, we can approximate any compact operator on a Hilbert space by the operators of finite rank.

Theorem 10.5.10. *Let X be a Hilbert space over \mathbf{C} . Then for each compact operator $T \in \mathcal{B}(X)$ there is a sequence of operators of finite rank converging to T in norm.*

Proof. First, assume X is separable. Then by Theorem 10.3.12 we see that X has a countable orthonormal basis $\{e_k\}_{k=1}^\infty$. For each $n \in \mathbf{N}$ define the operator P_n via

$$P_n(x) = \sum_{k=1}^n \langle x, e_k \rangle e_k \quad \text{for all } x \in X.$$

From Theorem 10.3.7 or Remark 10.4.20 it follows that

$$\max\{\|P_n\|, \|I - P_n\|\} \leq 1 \text{ for all } n \in \mathbf{N}$$

and

$$\lim_{n \rightarrow \infty} \|P_n(x) - x\| = 0.$$

Now, if the assertion were false, then there would be a compact operator $T \in \mathcal{B}(X)$ and a number $\delta > 0$ such that $\|T - F\| \geq \delta$ for all operators F of finite rank. Hence, for each $n \in \mathbf{N}$ there is an $x_n \in X$ such that $\|x_n\| = 1$ and $\|(T - F_n)x_n\| \geq 2^{-1}\delta$, where the operator F_n defined by

$$F_n(x) = P_n T(x) = \sum_{k=1}^n \langle T(x), e_k \rangle e_k \text{ for all } x \in X,$$

is an operator of finite rank. Since $\{x_n\}_{n=1}^\infty$ is bounded and T is compact, there is a subsequence $\{x_{n_j}\}_{j=1}^\infty$ such that $\{T(x_{n_j})\}_{j=1}^\infty$ converges to some element $w \in X$. But

$$\begin{aligned} \|(T - F_{n_j})(x_{n_j})\| &\leq \|(I - P_{n_j})(T(x_{n_j}) - w)\| + \|(I - P_{n_j})(w)\| \\ &\leq \|T(x_{n_j}) - w\| + \|(I - P_{n_j})(w)\|. \end{aligned}$$

Now for n_j big enough, we have

$$\|(I - P_{n_j})(w)\| < 2^{-2}\delta \text{ and } \|T(x_{n_j}) - w\| < 2^{-2}\delta.$$

The foregoing estimates therefore produce a contradiction: $2^{-1}\delta > 2^{-1}\delta$, which naturally proves the assertion in the case when X is separable.

Next, assume that X is any Hilbert space over \mathbf{C} . Set $S = T^*T$. Then S is self-adjoint and compact. By Theorem 10.5.8, S can be written as the following form

$$S(x) = \sum_{k=1}^{\infty} \lambda_k \langle x, e_k \rangle e_k, \quad x \in X,$$

for some orthonormal sequence $\{e_k\}_{k=1}^\infty$. Suppose now that X_0 is the subspace of all linear combinations of elements of the form $T^n(e_k)$ under $n+1, k \in \mathbf{N}$. Then $\overline{X_0}$ is a separable closed linear subspace of X , and T maps $\overline{X_0}$ into itself. By the first part, there exists a sequence $\{F_n\}_{n=1}^\infty$ of finite rank operators on $\overline{X_0}$ converging in norm to the restriction \overline{T} of T to $\overline{X_0}$. Now every element $x \in X$ can be decomposed into the form $x = y + z$ where $y \in \overline{X_0}$ and $z \in \overline{X_0}^\perp$.

is orthogonal to X_0 . With this, we get that z is orthogonal to each e_k , and so $z \in N(S)$. Accordingly,

$$0 = \langle S(z), z \rangle = \langle T^*T(z), z \rangle = \|T(z)\|^2;$$

that is, $T(z) = 0$. Thus, if $G_n(x) = F_n(y)$ for $n \in \mathbf{N}$, then G_n is of finite rank and $\lim_{n \rightarrow \infty} \|G_n - T\| = 0$ thanks to

$$\|G_n(x) - T(x)\| = \|F_n(y) - T(y)\| = \|F_n(y) - \check{T}(y)\| \leq \|F_n - \check{T}\| \|x\|.$$

□

To close this section, we use Remark 10.5.9 to produce the forthcoming Fredholm's alternative which arises in the study of the following homogeneous and inhomogeneous Fredholm integral equations:

$$f(x) - \int_{[a,b]} \kappa(x,y) f(y) dm_{id}(y) = 0; \quad f(x) - \int_{[a,b]} \kappa(x,y) f(y) dm_{id}(y) = g(x)$$

for a given pair $\{g, \kappa\}$ with:

$$g \in LRS_{id}^2([a,b], \mathbf{C}) \quad \text{and} \quad |\kappa(\cdot, \cdot)|^2 \in LRS_{id}([a,b] \times [a,b]).$$

Theorem 10.5.11. *Let X be a Hilbert space over \mathbf{C} and let T be a self-adjoint compact linear operator on X .*

(i) *If the only solution to the homogeneous equation $f - T(f) = 0$ is $f = 0$, then the inhomogeneous equation $f - T(f) = g$ has a unique solution f for each $g \in X$.*

(ii) *If $f - T(f) = 0$ has nonzero solutions, then $f - T(f) = g$ has a solution for a given $g \in X$ only if $\langle g, f \rangle = 0$ for any solution f to $f - T(f) = 0$, in which case $f - T(f) = g$ has infinitely many solutions, the difference of any two of them being a solution of $f - T(f) = 0$.*

Proof. (i) From Remark 10.5.9 it follows that X has an orthonormal basis $\{e_j\}_{j=1}^{\infty}$ comprising eigenvectors of T with eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$. For $g \in X$, let $g = \sum_{j=1}^{\infty} c_j e_j$, and then search a solution to $f - T(f) = g$ in the form $f = \sum_{j=1}^{\infty} d_j e_j$. Then

$$\sum_{j=1}^{\infty} d_j e_j - \sum_{j=1}^{\infty} c_j \lambda_j e_j = \sum_{j=1}^{\infty} c_j e_j$$

and hence $d_j = c_j(1 - \lambda_j)^{-1}$ provided $\lambda_j \neq 1$. Since $f - T(f) = 0$ has no nonzero solution, 1 is not an eigenvalue of T and consequently all d_j exist. This indicates that the solution to $f - T(f) = g$ must be of the form $\sum_{j=1}^{\infty} c_j(1 - \lambda_j)^{-1}e_j$. Of course, this shows that if $f - T(f) = g$ has a solution then it is unique. Now, it remains to check that the last series converges in X but also produces a solution to $f - T(f) = g$.

If 1 were a cluster point of the set of eigenvalues of T , then $\sum_{j=1}^{\infty} c_j(1 - \lambda_j)^{-1}e_j$ would not generally converge because its terms would be very large as λ_j was close to 1. But, we know from Remarks 10.5.7 (ii) that $\lambda_j \rightarrow 0$ as $j \rightarrow \infty$, and so no subsequence of $\{\lambda_j\}$ can approach 1 – this means that 1 is not a cluster point of the set $\{\lambda_j\}$ and so that $\sup_{j \in \mathbb{N}} |1 - \lambda_j|^{-2} < \infty$. Consequently,

$$\sum_{j=1}^{\infty} |c_j(1 - \lambda_j)^{-1}|^2 \leq \left(\sup_{j \in \mathbb{N}} |1 - \lambda_j|^{-2} \right) \sum_{j=1}^{\infty} |c_j|^2 < \infty.$$

This finiteness, along with Theorem 10.3.10, implies that $\sum_{j=1}^{\infty} c_j(1 - \lambda_j)^{-1}e_j$ is convergent in X .

Next, let us verify that the last series is a solution to $f - T(f) = g$. Since $g \in X$ and $\{e_j\}$ is an orthonormal basis for X , we conclude that $\sum_{j=1}^{\infty} |\langle g, e_j \rangle|^2 < \infty$. Also since T is self-adjoint and compact, it follows that $\lambda = \inf_k |1 - \lambda_k| > 0$ and then

$$\sum_{j=1}^{\infty} |\langle g, e_j \rangle|^2 |1 - \lambda_j|^{-2} \leq \lambda^{-2} \sum_{j=1}^{\infty} |\langle g, e_j \rangle|^2 < \infty.$$

This yields that $f = \sum_{j=1}^{\infty} \langle g, e_j \rangle (1 - \lambda_j)^{-1} e_j$ is convergent in X thanks to Theorem 10.3.10. Accordingly,

$$\begin{aligned} T(f) &= \sum_{j=1}^{\infty} \langle g, e_j \rangle (1 - \lambda_j)^{-1} T(e_j) \\ &= \sum_{j=1}^{\infty} \langle g, e_j \rangle (1 - \lambda_j)^{-1} \lambda_j e_j \\ &= \sum_{j=1}^{\infty} c_j (1 - \lambda_j)^{-1} e_j - \sum_{j=1}^{\infty} \langle g, e_j \rangle \lambda_j e_j \\ &= f - g. \end{aligned}$$

(ii) Given $g \in X$, assume now that f_1 satisfies $f - T(f) = g$ and f_2 is a nonzero solution to $f - T(f) = 0$. This assumption, plus $T^* = T$, gives

$$\langle f_1, f_2 \rangle = \langle T(f_1), f_2 \rangle + \langle g, f_2 \rangle = \langle f_1, T(f_2) \rangle + \langle g, f_2 \rangle = \langle f_1, f_2 \rangle + \langle g, f_2 \rangle,$$

whence deriving $\langle g, f_2 \rangle = 0$. Here, since $f - T(f) = g$ implies $(f + cf_2) - T(f + cf_2) = g$ for any f_2 with $f_2 - T(f_2) = 0$ and $c \in \mathbf{C}$, there are infinitely many such solutions to $f - T(f) = g$. The final assertion follows simply from taking the difference between two solutions to $f - T(f) = g$. \square

Clearly, the above Fredholm's alternative theorem is a simple generalization of facts on sets of linear algebraic equations: An inhomogeneous system has a unique solution when and only when the determinant of the coefficients is nonzero, i.e., the corresponding homogeneous system has no nonzero solution. In addition, we here want to say by an example that the compactness of the operator T in the Fredholm's alternative theorem is essential.

Example 10.5.12. On $LRS_{id}^2([0, 1], \mathbf{C})$ define $T(f)(x) = (1 - x)f(x)$. Then T is self-adjoint but not compact. The homogeneous equation $f - T(f) = 0$ has no solution other than $f = 0$. Nevertheless, the inhomogeneous equation $f - T(f) = g$ has no solution in $LRS_{id}^2([0, 1], \mathbf{C})$ for any nonconstant function $g \in LRS_{id}^2([0, 1], \mathbf{C})$ – indeed, the solution to $f - T(f) = g$ is then proportional to x^{-1} as $x \in (0, 1]$, which is certainly not in $LRS_{id}^2([0, 1], \mathbf{C})$.

Problems

10.1. Equip $C([-1, 1], \mathbf{C})$ with the 2-norm and the inner product $\langle f, g \rangle = \int_{-1}^1 f(x)g(x)dx$. Prove that $C([-1, 1], \mathbf{C})$ is an inner product space but not a Hilbert space.

10.2. Equip $\ell_p(\mathbf{N}, \mathbf{C})$, $1 \leq p < \infty$, with the p -norm. Prove that $\ell_2(\mathbf{N}, \mathbf{C})$ is a Hilbert space under the inner product

$$\langle x, y \rangle = \sum_{j=1}^{\infty} x_j \overline{y_j} \text{ for all } x = (x_j)_{j=1}^{\infty}, y = (y_j)_{j=1}^{\infty} \in \ell_2(\mathbf{N}, \mathbf{C}),$$

but $\ell_p(\mathbf{N}, \mathbf{C})$ is not a Hilbert space under the p -norm whenever $p \neq 2$.

10.3. Let X be an inner product space over \mathbf{F} .

(i) If $\mathbf{F} = \mathbf{R}$, prove that $x \perp y$ is equivalent to $\|x + y\|^2 = \|x\|^2 + \|y\|^2$. Is this equivalence true for $\mathbf{F} = \mathbf{C}$?

(ii) $x \perp y$ if and only if $\|x + \beta y\| \geq \|x\|$ for all $\beta \in \mathbf{F}$.

10.4. Let X be a Hilbert space over \mathbf{F} . Prove that if M is a closed linear subspace of X then $(M^\perp)^\perp = M$.

10.5. Equip $LRS_{id}^2([-1, 1], \mathbf{C})$ with the inner product $\langle f, g \rangle = \int_{[-1, 1]} f \bar{g} dm_{id}$.

(i) Let

$$M = \{f \in LRS_{id}^2([-1, 1], \mathbf{C}) : f(x) = 0 \text{ for all } x \in [-1, 0]\}.$$

Find M^\perp .

(ii) Let

$$\begin{cases} M_{odd} = \{f \in LRS_{id}^2([-1, 1], \mathbf{C}) : f(-x) = -f(x) \text{ for all } x \in [-1, 1]\}; \\ M_{even} = \{f \in LRS_{id}^2([-1, 1], \mathbf{C}) : f(-x) = f(x) \text{ for all } x \in [-1, 1]\}. \end{cases}$$

Prove $LRS_{id}^2([-1, 1], \mathbf{C}) = M_{odd} \oplus M_{even}$.

10.6. Equip $LRS_{id}^2([- \pi, \pi], \mathbf{C})$ with the inner product $\langle f, g \rangle = \int_{[- \pi, \pi]} f \bar{g} dm_{id}$. Prove the following two results.

(i) $\{(2\pi)^{-\frac{1}{2}} e^{inx}\}_{n=-\infty}^{\infty}$ is an orthonormal set.

(ii)

$$\{(2\pi)^{-\frac{1}{2}}, \pi^{-\frac{1}{2}} \cos t, \pi^{-\frac{1}{2}} \sin t, \pi^{-\frac{1}{2}} \cos 2t, \pi^{-\frac{1}{2}} \sin 2t, \dots\}$$

is an orthonormal basis of $LRS_{id}^2([- \pi, \pi], \mathbf{C})$.

10.7. Let $LRS_{id}^2([- \pi, \pi], \mathbf{C})$ be as in Problem 10.5. For $\phi \in C([- \pi, \pi], \mathbf{C})$ let $T(f) = \phi f$ be defined on $LRS_{id}^2([- \pi, \pi], \mathbf{C})$.

(i) Calculate T^* using the inner product defined above.

(ii) Prove that if ϕ is real-valued then T is self-adjoint.

(iii) Find a condition on ϕ such that T is respectively unitary, nonnegative, or a projection.

10.8. Let X be a Hilbert space over \mathbf{C} and $T \in \mathcal{B}(X)$. Prove the following results.

(i) $R(T)^\perp = N(T^*)$ and $N(T)^\perp = \overline{R(T^*)}$.

(ii) $R(T^*)^\perp = N(T)$ and $N(T^*)^\perp = \overline{R(T)}$.

10.9. Equip $C([0, 2\pi], \mathbf{C})$ with the sup-norm: $\|f\|_\infty = \sup_{x \in [0, 2\pi]} |f(x)| < \infty$.

- (i) Define $T(f)(x) = xf(x)$ for $f \in C([0, 2\pi], \mathbf{C})$. Find $\sigma(T)$.
- (ii) Define $T(f)(x) = e^{ix}f(x)$ for $f \in C([0, 2\pi], \mathbf{C})$. Prove $\sigma(T) = \{\lambda \in \mathbf{C} : |\lambda| = 1\}$.

10.10. Equip $\ell_2(\mathbf{N}, \mathbf{C})$ with the 2-norm.

- (i) If $T(x) = T((x_1, x_2, \dots)) = (x_2, x_3, \dots)$, then find $\sigma(T)$.
- (ii) If $T(x) = T((x_1, x_2, \dots)) = (0, -x_1, -x_2, -x_3, \dots)$, then prove the following results:
- (a) T has no eigenvalue;
 - (b) $\rho(T) = \{\lambda \in \mathbf{C} : |\lambda| > 1\}$;
 - (c) $\|(T - \lambda I)^{-1}\| = (|\lambda| - 1)^{-1}$.

10.11. Suppose that X is a Hilbert space over \mathbf{C} , and that $T \in \mathcal{B}(X)$. Prove $\sigma(T^*) = \{\bar{\lambda} : \lambda \in \sigma(T)\}$.

10.12. Let $(X, \|\cdot\|)$ be a Hilbert space over \mathbf{C} . If $T \in \mathcal{B}(X)$ is compact and self-adjoint and $x \in X$ satisfies $\|x\| = 1$, then prove the following two results.

- (i) $\lambda \in \mathbf{R}$ obeys $\|T(x) - \lambda x\|^2 = \|T(x)\|^2 - \langle T(x), x \rangle^2$ when and only when $\lambda = \langle T(x), x \rangle$.
- (ii) There exists a $\lambda \in \sigma(T)$ such that $|\lambda - \langle T(x), x \rangle| \leq \sqrt{\|T(x)\|^2 - \langle T(x), x \rangle^2}$.

10.13. Let $T(f)(x) = \int_0^1 e^{xy} f(y) dy$. Find the eigenvalues and eigenvectors of T .

10.14. Let X be a Hilbert space over \mathbf{C} . If $T = T_1 + iT_2$ is normal operator on X , then prove $\|T\|^2 = \|T_1^2 + T_2^2\|$ and $\|T^2\| = \|T\|^2$.

10.15. Let X be a Hilbert space over \mathbf{C} , $\lambda \in \mathbf{C}$, and $T \in \mathcal{B}(X)$ normal. Prove the following results:

- (i) $\|(T^* - \bar{\lambda}I)(x)\| = \|(T - \lambda I)(x)\|$ for $x \in X$;
- (ii) $T(x) = \lambda x \Rightarrow T^*(x) = \bar{\lambda}x$;
- (iii) $r_\sigma(T) = \|T\|$.

10.16. Let X be a Hilbert space over \mathbf{C} . Prove if $T \in \mathcal{B}(X)$ and T^*T is compact then T is compact.

10.17. Suppose that $[a_{mn}]$ is an infinite matrix of real numbers obeying $\sum_{m,n=1}^{\infty} a_{mn}^2 < \infty$. Let the linear operator T on ℓ_2 be defined by $T(x) = y$ where $x = (x_1, x_2, \dots)$ and $y = (\sum_{n=1}^{\infty} a_{1n}x_n, \sum_{n=1}^{\infty} a_{2n}x_n, \dots)$. Prove that T is a compact operator on ℓ_2 .

10.18. Equip the real-valued Lebesgue space $LRS_{id}^2[0, 1]$ with the inner product $\langle f, g \rangle = \int_{[0,1]} f \bar{g} dm_{id}$. Suppose that

$$K(f)(x) = \int_{[0,1]} \kappa(x, y) f(y) dy \text{ for all } f \in LRS_{id}^2[0, 1].$$

(i) If

$$\kappa(y, x) = \sum_{k=1}^n g_k(x) f_k(y) \text{ where } f_k, g_k \in LRS_{id}^2[0, 1]$$

and $\{g_k\}_{k=1}^n$ are linearly independent, then prove that T is compact.

(ii) If λ is nonzero eigenvalue, prove its corresponding eigenvector has the following form:

$$f(x) = \sum_{k=1}^n c_k g_k(x) \text{ where } c_k \text{ is constant;}$$

moreover, if $h_{k,j} = \int_{[0,1]} f_k g_j dm_{id}$, then $c_k = \lambda^{-1} \sum_{j=1}^n h_{j,k} c_j$ for $k = 1, 2, \dots, n$.

(iii) If $f_k = g_k$ and $\langle f_j, g_k \rangle = 0$ for $j \neq k$, evaluate eigenvalues and eigenvectors of K .

(iv) If $\kappa(y, x) = \cos \pi^{-1}(y + x)$ for $x, y \in [0, 1]$, evaluate eigenvalues and eigenvectors of K .

10.19. Suppose that T is a compact operator on an infinite-dimensional Hilbert space $(X, \|\cdot\|)$ over \mathbf{C} . If $\{e_j\}_{j=1}^{\infty}$ is an orthonormal basis of X , then prove $\lim_{j \rightarrow \infty} \|T(e_j)\| = 0$.

10.20. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Hilbert spaces over \mathbf{C} . A bounded linear operator $T : X \rightarrow Y$ is called Hilbert-Schmidt (Erhard Schmidt) operator provided that there exists an orthonormal basis $\{e_j\}_{j=1}^{\infty}$ in X such that $\sum_{j=1}^{\infty} \|T(e_j)\|_Y^2 < \infty$.

(i) If $T^* : Y \rightarrow X$ is defined by

$$\langle T(x), y \rangle_Y = \langle x, T^*(y) \rangle_X \text{ for all } (x, y) \in X \times Y,$$

where $\langle \cdot, \cdot \rangle_X$ and $\langle \cdot, \cdot \rangle_Y$ stand for the inner products equipped with X and Y respectively, prove that T is Hilbert-Schmidt operator when and only when T^* is Hilbert-Schmidt operator.

(ii) Prove that any Hilbert-Schmidt operator is compact.

(iii) If $\kappa(\cdot, \cdot) : (a, b) \times (c, d) \rightarrow \mathbf{C}$ satisfies $\int_{(a,b)} \int_{(c,d)} |\kappa(x, y)|^2 dm_{id}(x) dm_{id}(y) < \infty$ where $(a, b) \times (c, d) \subseteq \mathbf{R}^2$, then prove that the integral operator

$$T(f)(x) = \int_{(c,d)} \kappa(x, y) f(y) dm_{id}(y) \text{ for all } x \in (a, b)$$

is a Hilbert-Schmidt operator from $LRS_{id}^2((c, d), \mathbf{C})$ to $LRS_{id}^2((a, b), \mathbf{C})$ and therefore compact.

10.21. Let X and Y be Hilbert spaces over \mathbf{C} . Suppose $\{e_j\}_{j=1}^\infty$ and $\{f_k\}_{k=1}^\infty$ are orthonormal bases of X and Y respectively. For a sequence $\{c_k\}_{k=1}^\infty$ in \mathbf{C} define a linear operator $T : X \rightarrow Y$ by

$$T(x) = \sum_{k=1}^{\infty} c_k \langle x, e_k \rangle_X f_k \text{ for all } x \in X,$$

where $\langle \cdot, \cdot \rangle_X$ and $\langle \cdot, \cdot \rangle_Y$ stand for the inner products equipped with X and Y respectively. Prove the following results:

(i) T is a bounded operator when and only when $\sup_{k \in \mathbf{N}} |c_k| < \infty$;

(ii) T is a compact operator when and only when $\lim_{k \rightarrow \infty} |c_k| = 0$;

(iii) T is a Hilbert-Schmidt operator when and only when $\sum_{k=1}^{\infty} |c_k|^2 < \infty$;

(iv) T is an operator of finite rank when and only when there is an $N \in \mathbf{N}$ such that $|c_n| = 0$ for $n > N$.

Hints or Solutions

This part comprises hints or solutions to those problems included at the end of each chapter in the text.

1 Preliminaries

1.1 If $x \in (X \cup Y) \setminus (X \cap Y)$ then $x \in X \cup Y$ but $x \notin X \cap Y$ and hence we must have: $x \in X$ but $x \notin Y$ or $x \in Y$ but $x \notin X$, that is, $x \in (X \setminus Y) \cup (Y \setminus X)$. Thus $(X \cup Y) \setminus (X \cap Y) \subseteq (X \setminus Y) \cup (Y \setminus X)$. The reversed inclusion can be proved similarly. Therefore, the desired equality follows.

1.2 $\cup_{j \in \mathbf{N}} X_j = [0, 1]$ and $\cap_{j \in \mathbf{N}} X_j = \{0\}$.

1.3 If $x \in X \cap (\cup_{j \in I} X_j)$ then $x \in X$ and $x \in X_j$ for some $j \in I$ and hence $x \in \cup_{j \in I} (X \cap X_j)$. Consequently, $X \cap (\cup_{j \in I} X_j) \subseteq \cup_{j \in I} (X \cap X_j)$. The reversed inclusion can be proved similarly. So, the desired equality follows. The second equality can be checked in a similar manner.

1.4 This follows from the related definitions.

1.5 $f(x) = a + (b - a)x: (0, 1) \rightarrow (a, b)$ and $g(x) = \tan 2^{-1}\pi x: (0, 1) \rightarrow (0, \infty)$ are bijective.

1.6 If $x \in (g \circ f)^{-1}(C)$ then $g \circ f(x) \in C$ and hence $f(x) \in g^{-1}(C)$ and $x \in f^{-1}(g^{-1}(C))$. Accordingly, $(g \circ f)^{-1}(C) \subseteq f^{-1}(g^{-1}(C))$. Similarly, we have $(g \circ f)^{-1}(C) \supseteq f^{-1}(g^{-1}(C))$, thereby getting the desired equality.

1.7 Since X is infinite, $X \setminus \{x\}$ is infinite and hence there is an $x_1 \in X \setminus \{x\}$. Of course, $X \setminus \{x, x_1\}$ is infinite, and then there is an $x_2 \in X \setminus \{x, x_1\}$. This process produces a sequence of distinct points $\{x_j\}_{j=1}^{\infty}$ in $X \setminus \{x\}$. The map: $x \rightarrow x_1, x_j \rightarrow x_{j+1}; j \in \mathbf{N}, y \rightarrow y$ otherwise, is a bijection of X onto $X \setminus \{x\}$.

1.8 Since $f(X)$ is a subset of Y , $f(X)$ is countable. While f is one-to-one, X is countable too.

1.9 If there is a function ϕ from $[0, 1]$ onto X , then for each $x \in [0, 1]$ we have $\phi_x = \phi(x) \in X$ which is a function from $[0, 1]$ to \mathbf{R} . For any $x \in [0, 1]$ let

$$f(x) = \begin{cases} 0, & \phi_x(x) \neq 0 \\ 1, & \phi_x(x) = 0 \end{cases}.$$

Since ϕ is surjective, there is an $x_0 \in [0, 1]$ such that $f = \phi_{x_0}$. But $f(x_0) \neq \phi_{x_0}(x_0)$, a contradiction. Therefore, no such ϕ exists.

1.10 (i) $a^{-1}ax = a^{-1}b$ gives $x = a^{-1}b$ and conversely.

(ii) Use $a^{-1}(a^{-1})^{-1} = 1$.

(iii) Use $(ab)(ab)^{-1} = 1$ and $aba^{-1}b^{-1} = 1$.

1.11 Obviously, $|a_1| \leq |a_1|$. Assuming the inequality for $n = k$, we get by the triangle inequality that

$$\left| \sum_{j=1}^{k+1} a_j \right| \leq \left| \sum_{j=1}^l a_j \right| + |a_{k+1}| \leq \sum_{j=1}^k |a_j| + |a_{k+1}|,$$

as desired.

1.12 Use $b^3 - a^3 = (b - a)(b^2 + ba + a^2)$ and $b^2 + ba + a^2 \geq 0$ (under $a, b \geq 0$).

1.13 Clearly, $\sup(X + Y) \leq \sup X + \sup Y$. For any $\varepsilon > 0$ there are $x \in X$ and $y \in Y$ such that $\sup X < x + \varepsilon$ and $\sup Y < y + \varepsilon$. So, $\sup X + \sup Y < \sup(X + Y) + 2\varepsilon$. This gives $\sup X + \sup Y \leq \sup(X + Y)$. Thus, $\sup(X + Y) = \sup X + \sup Y$. The proof of $\inf(X + Y) = \inf X + \inf Y$ is similar to the previous one.

1.14 (i) Since $(a_n)_{n=1}^{\infty}$ is increasing and bounded from above by any b_k , $a = \lim_{n \rightarrow \infty} a_n$ exists and belongs to $\cap_{n \in \mathbf{N}} I_n$. Similarly, $b = \lim_{n \rightarrow \infty} b_n$ exists and belongs to $\cap_{n \in \mathbf{N}} I_n$.

(ii) This follows from (i).

(iii) This follows from (i) and (ii).

1.15 Use Definition 1.2.6 and some simple estimates.

1.16 Use $s_n > a > 0$ if and only if $0 < s_n^{-1} < a^{-1}$.

1.17 Use $s_m - s_n = (n + 1)^{-1} + \cdots + m^{-1}$ for $m > n$.

1.18 (i) It suffices to prove the first equality. Since $\{s_n\}_{n \in \mathbf{N}}$ is bounded, $a = \limsup s_n \in \mathbf{R}$ and for any $\varepsilon > 0$ there is an $n_0 \in \mathbf{N}$ such that $n > n_0$ implies $s_n < a + \varepsilon$. Thus for all $m > n_0$ we have $\sup\{s_n : n > m\} \leq a + \varepsilon$. Of course, this yields $\sup\{s_n : n > m\} \leq a$ as $m > n_0$ and then $\lim_{m \rightarrow \infty} \sup\{s_n : n > m\} \leq a$. For the reversed inequality, let $m \in \mathbf{N}$ be fixed. Then for any $\varepsilon > 0$ there exists $n_0 > m$ such that $s_{n_0} > a - \varepsilon$. Hence $\sup\{s_n : n > m\} \geq a$. This is valid for any $m \in \mathbf{N}$. So, $a \leq \lim_{m \rightarrow \infty} \sup\{s_n : n > m\}$.

(ii) If $\lim_{n \rightarrow \infty} s_n = s \in \mathbf{R}$ exists, then all subsequences of $\{s_n\}_{n \in \mathbf{N}}$ have limit s . Thus $S = \{s\}$ and consequently $\limsup s_n = s = \liminf s_n$. Conversely, assume $\limsup s_n = s = \liminf s_n$. If $\{s_n\}_{n=1}^\infty$ has no limit s , then there is an $\varepsilon_0 > 0$ and a subsequence $(s_{n_k})_{k \in \mathbf{N}}$ of $\{s_n\}_{n=1}^\infty$ such that $|s_{n_k} - s| \geq \varepsilon_0$. Since $\{s_{n_k}\}_{k=1}^\infty$ is bounded, the Bolzano-Weierstrass property implies that this subsequence has a subsequence which converges to $t \in \mathbf{R}$. Of course, $|t - s| \geq \varepsilon_0$. This indicates $t \in S$ and so $\sup S$ and $\inf S$ cannot both be s , contradicting the assumption.

2 Riemann Integrals

2.1 (i) Given $\varepsilon > 0$, choose $\delta = 2^{-1}\varepsilon$. If P is any partition of $[0, 1]$, then in $S(f, P, \xi)$ there are at most two nonzero terms (2^{-1} could be a partition point, say $x_j = 2^{-1}$, and we could have $\xi_j = x_j = \xi_{j+1}$). Thus $\|P\| < \delta$ implies $0 \leq S(f, P, \xi) \leq 2\|P\| < \varepsilon$.

(ii) Given $\varepsilon > 0$, there are at most $2\varepsilon^{-1}$ of the numbers k^{-1} that satisfy $k^{-1} > 2^{-1}\varepsilon$. Thus in any sum $P(f, \xi)$, those numbers contribute at most $(2/\varepsilon)2(\sup_{x \in [0, 1]} |f(x)|)\|P\| = 4\varepsilon^{-1}\|P\|$. The remaining terms in $S(f, P, \xi)$ contribute at most $\sup_{x \in [0, 1]} |f(x)|$ times the total length of those subintervals in $[0, 2^{-1}\varepsilon]$. Therefore $0 \leq S(f, P, \xi) \leq (4\varepsilon^{-1})\|P\| + (2^{-1}\varepsilon)$. Define $\delta = 2^{-3}\varepsilon^2$ to get $|S(f, P, \xi)| < \varepsilon$ whenever $\|P\| < \delta$.

(iii) Let P be a partition of $[0, 2]$ and suppose $x_j \leq 1 \leq x_{j+1}$. In $S(f, P, \xi)$, if $k < j$ then $f(\xi_k) = 1$; and if $k > j + 1$ then $f(\xi_k) = -2$. Thus, if $\Delta x_k = x_k - x_{k-1}$, then

$$\begin{aligned} S(f, P, \xi) &= \sum_{k=1}^{j-1} \Delta x_k + f(\xi_j) \Delta x_j + f(\xi_{j+1}) \Delta x_{j+1} - 2 \sum_{k=j+2}^n \Delta x_k \\ &= x_{j-1} + f(\xi_j) \Delta x_j + f(\xi_{j+1}) \Delta x_{j+1} - 2(2 - x_{j+1}). \end{aligned}$$

Given $\varepsilon > 0$, choose $\delta = 2^{-2}\varepsilon$. Then for $\|P\| < \delta$, one has

$$\begin{aligned} |S(f, P, \xi) + 1| &= |x_{j-1} - 1 + f(\xi_j)\Delta x_j + f(\xi_{j+1})\Delta x_{j+1} - 2(1 - x_{j+1})| \\ &\leq \|P\| + \|P\| + 2\|P\| = 4\|P\| < \varepsilon. \end{aligned}$$

(iv) Let P be a partition of $[0, 2]$ and suppose $x_j \leq 1 \leq x_{j+1}$. In $S(f, P, \xi)$, if $k < j$ then $f(\xi_k) = 2$; and if $k > j + 1$ then $f(\xi_k) = 1$. Thus, if $\Delta x_k = x_k - x_{k-1}$, then

$$\begin{aligned} S(f, P, \xi) &= 2 \sum_{k=1}^{j-1} \Delta x_k + f(\xi_j)\Delta x_j + f(\xi_{j+1})\Delta x_{j+1} + \sum_{k=j+2}^n \Delta x_k \\ &= 2x_{j-1} + f(\xi_j)\Delta x_j + f(\xi_{j+1})\Delta x_{j+1} + (2 - x_{j+1}). \end{aligned}$$

Given $\varepsilon > 0$, choose $\delta = \varepsilon/(14)$. Then for $\|P\| < \delta$, one has

$$\begin{aligned} |S(f, P, \xi) - 3| &= |2x_{j-1} + f(\xi_j)\Delta x_j + f(\xi_{j+1})\Delta x_{j+1} + 2 - x_{j+1} - 3| \\ &\leq |f(\xi_j)\Delta x_j| + |f(\xi_{j+1})\Delta x_{j+1}| + |x_{j-1} - 1| + |x_{j-1} - x_{j+1}| \\ &\leq 5\|P\| + 5\|P\| + 2\|P\| + 2\|P\| = 14\|P\| < \varepsilon. \end{aligned}$$

2.2 Let $P = \{0, 1/n, 2/n, \dots, 1\}$ and $\mu_k = k/n$; then $S(f, P, \xi) = n^{-1} \sum_{k=1}^n f(k/n)$. And by the geometric definition of $\int_0^1 f(x) dx$, one has $\int_0^1 \sqrt{1-x^2} dx = \pi/4$.

2.3 f is integrable on $[-1, 0]$; but f is unbounded on both $[0, 1]$ and $[-1, 1]$, so it is not integrable there.

2.4 This follows immediately from Corollary 2.2.2

2.5 Since f is increasing,

$$m_k = f(x_{k-1}) = \frac{k-1}{n} \quad \text{and} \quad L(f, P_n) = \sum_{k=1}^n \frac{k-1}{n} \left(\frac{k}{n} - \frac{k-1}{n} \right) = \frac{n-1}{2n}.$$

Similarly,

$$M_k = f(x_k) = \frac{k}{n} \quad \text{and} \quad U(f, P_n) = \sum_{k=1}^n \frac{k}{n} \left(\frac{k}{n} - \frac{k-1}{n} \right) = \frac{n+1}{2n}.$$

Thus $U(f, P_n) - L(f, P_n) = n^{-1}$, and it is clear that for any $\varepsilon > 0$ we can choose $n > \varepsilon^{-1}$ and get $U(f, P_n) - L(f, P_n) < \varepsilon$. Therefore f is integrable on $[0, 1]$. Also $\int_0^1 f dx = \lim_{n \rightarrow \infty} U(f, P_n) = 2^{-1}$.

2.6 Define $f_+(x) = \max\{f(x), 0\}$ and $f_-(x) = -\min\{f(x), 0\}$. Then

$$f(x) = f_+(x) - f_-(x); \quad |f(x)| = f_+(x) + f_-(x).$$

So, $f \in R[a, b]$ implies $f_{\pm} \in R[a, b]$ and hence $|f| \in R[a, b]$ with

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b f_+(x) dx + \int_a^b f_-(x) dx = \int_a^b |f(x)| dx.$$

For the converse, taking $f(x) = 1$ when $x \in [a, b] \cap \mathbf{Q}$ and $f(x) = -1$ otherwise, one gets that $|f| = 1$ but $f \notin R[a, b]$.

2.7 Let $f \in R[a, b]$. Then for any partition P of $[c, d]$ then $P' = P \cup \{a, b\}$ is a partition of $[a, b]$, and $U(f, P) - L(f, P) \leq U(f, P') - L(f, P')$ and hence $f \in R[c, d]$.

Conversely, let $f \in R[c, d]$ for any $[c, d] \subset (a, b)$. Then for any $\varepsilon > 0$ we may take $[c, d]$ such that $a < c \leq a + \varepsilon/(6K)$ and $b - \varepsilon/(bK) \leq d < b$, where $K = \sup_{x \in [a, b]} |f(x)| + 1$. Because f is integrable on $[c, d]$, one can choose a partition P^* of $[c, d]$ such that $U(f, P^*) - L(f, P^*) < 3^{-1}\varepsilon$. Now let $P = \{a\} \cup P^* \cup \{b\}$. Then P is a partition of $[a, b]$, and

$$\begin{aligned} U(f, P) - L(f, P) &= (M_1 - m_1)(c - a) + U(f, P^*) - L(f, P^*) + (M_n - m_n)(b - d) \\ &\leq 2K(c - a) + U(f, P^*) - L(f, P^*) + 2K(b - d) \\ &< 2K(\varepsilon/(6K)) + (3^{-1}\varepsilon) + 2K(\varepsilon/(6K)) = \varepsilon. \end{aligned}$$

Hence $f \in R[a, b]$.

2.8 This follows the Cauchy-Schwarz inequality:

$$\int_0^\pi \sqrt{x \sin x} dx \leq \left(\int_0^\pi x dx \right)^{2^{-1}} \left(\int_0^\pi \sin x dx \right)^{2^{-1}} = \pi.$$

2.9 Since $(f + g)^2 = f^2 + 2fg + g^2$, by the Cauchy-Schwarz inequality

$$\begin{aligned} &\int_a^b (f(x) + g(x))^2 dx \\ &\leq \int_a^b f^2(x) dx + 2 \left(\int_a^b f^2(x) dx \right)^{2^{-1}} \left(\int_a^b g^2(x) dx \right)^{2^{-1}} + \int_a^b g^2(x) dx \\ &= \left(\left(\int_a^b f^2(x) dx \right)^{2^{-1}} + \left(\int_a^b g^2(x) dx \right)^{2^{-1}} \right)^2, \end{aligned}$$

as desired.

2.10 It is easy to see that if $f(x) \geq \delta > 0$ then $\inf(1/f) = 1/(\sup(f))$ and $\sup(1/f) = 1/(\inf(f))$ and hence

$$U(1/f, P) - L(1/f, P) = \sum_{k=1}^n \frac{M_{f,k} - m_{f,k}}{M_{f,k} m_{f,k}} (x_k - x_{k-1}) \leq \delta^{-2} (U(f, P) - L(f, P)).$$

This implies $1/f \in R[a, b]$.

2.11 Given $\varepsilon > 0$, choose $\delta = 2^{-2}\varepsilon^2$ and $x_1, x_2 \in [0, \infty)$. If both $x_1, x_2 < 2^{-2}\varepsilon^2$, then $|x_1 - x_2| < \delta$ and $|\sqrt{x_1} - \sqrt{x_2}| \leq \sqrt{x_1} + \sqrt{x_2} < \varepsilon$. If either $x_1 \geq 2^{-2}\varepsilon^2$ or $x_2 \geq 2^{-2}\varepsilon^2$, then $\sqrt{x_1} + \sqrt{x_2} \geq 2^{-1}\varepsilon$. Thus

$$|x_1 - x_2| < \delta \Rightarrow |\sqrt{x_1} - \sqrt{x_2}| = |x_1 - x_2|(\sqrt{x_1} + \sqrt{x_2})^{-1} < \varepsilon.$$

We note that an alternate proof exists using $\sqrt{|x_1 - x_2|} \leq \sqrt{x_1} + \sqrt{x_2}$ and the choice of $\delta = \varepsilon^2$.

2.12 Suppose that f is unbounded on (a, b) . For each $n \in \mathbf{N}$ choose $s_n \in (a, b)$ such that $|f(s_n)| \geq n$. Then $\{s_n\}$ has a convergent subsequence $\{s_{n(k)}\}$, which means that $\{s_{n(k)}\}$ is a Cauchy sequence in (a, b) , while the unbounded sequence $\{f(s_{n(k)})\}$ is not a Cauchy sequence. Hence f is not uniformly continuous on (a, b) .

2.13 Taking $x_{1,n} = (2n\pi)^{-1}$ and $x_{2,n} = (2n\pi + 2^{-1}\pi)^{-1}$, we have $|x_{1,n} - x_{2,n}| \rightarrow 0$ as $n \rightarrow \infty$ but $|f(x_{1,n}) - f(x_{2,n})| = 1 > \varepsilon$ for any small enough number $\varepsilon > 0$.

2.14 The function f is piecewise linear and is therefore continuous and bounded on $[0, k - 2^{-1}] \setminus \{j\}_{j=1}^{k-1}$. By defining a continuous function g on $[0, k - 2^{-1}]$ and $g = f$ on $[0, k - 2^{-1}] \setminus \{j\}_{j=1}^{k-1}$, we see that f is in $R[0, k - 2^{-1}]$.

2.15 Define $g(x) = x^{-1} \sin x$ or $g(x) = 1$ if $x \neq 0$ or $x = 0$. Then g is continuous on $[0, 5]$ and so $g \in R[0, 5]$. Consequently, $f \in R[0, 5]$.

2.16 Since $f \in R[a, b]$, f is bounded and $F(x)$ exists for any $x \in [a, b]$. The continuity of F follows from

$$|F(x) - F(x_0)| = \left| \int_{x_0}^x f(t) dt \right| \leq \sup_{x \in [a, b]} |f(x)| |x - x_0|.$$

2.17 If $x_1 < x_2$, then there is a $\mu \in [a, b]$ such that

$$F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t) dt \geq \min_{x \in [a, b]} f(x) (x_2 - x_1) = f(\mu) (x_2 - x_1) > 0$$

since $f \in C[a, b]$ and $f > 0$.

2.18 Because $f(2/(\pi k)) = \sin(\pi k/2)$ has no limit as $k \rightarrow \infty$, f is not continuous at 0, and of course not continuous on $[0, 1/\pi]$. But f is bounded on $[0, 1/\pi]$ and continuous on $[\varepsilon, 1/\pi]$ for every $\varepsilon > 0$. So $f \in R[0, 1/\pi]$ by Problem 2.7.

2.19 $\int_1^\infty x^p dx$ is convergent if and only if $p < -1$ and $\int_0^1 x^p dx$ is convergent if and only if $p > -1$.

2.20 (i) convergent; (ii) convergent; (iii) divergent; (iv) divergent: $(\frac{1}{\ln x})' = -\frac{1}{x(\ln x)^2}$, and $\lim_{t \rightarrow 1^+} \frac{-1}{\ln t} = -\infty$. This implies that $\int_{1^+}^2 \frac{x}{(\ln x)^2} dx$ is divergent. Note that $\int_2^\infty \frac{x}{(\ln x)^2} dx$ is convergent, but this not sufficient to infer convergence over $[1, \infty)$.

2.21 Use $(\ln x)' = 1/x$ for $x > 0$.

2.22 Let $F(x) = \int_x^\infty e^{-a|x-y|-b|y|} dy$. Then $f * g(x) = F(x) + F(-x)$ and it suffices to evaluate $F(x)$. If $x \geq 0$ then $F(x) = \frac{1}{(a+b)e^{bx}}$ and if $x < 0$ then we have two subcases: Firstly, $a \neq b \Rightarrow F(x) = \frac{e^{ax}}{a+b} + \frac{e^{ax}-e^{bx}}{b-a}$; Secondly, $a = b \Rightarrow F(x) = -xe^{ax} + \frac{e^{ax}}{2a}$.

2.23 Since the polynomials are dense in $C[0, 1]$, $\int_0^1 f(x)p(x) dx = 0$ for any polynomial p and consequently, for any $\varepsilon > 0$ there is a polynomial p such that $\max_{x \in [0, 1]} |f(x) - p(x)| < \varepsilon$ and hence

$$\int_0^1 f^2(x) dx = \int_0^1 (f(x) - p(x))f(x) dx \leq \varepsilon \int_0^1 |f(x)| dx$$

implying $\int_0^1 f^2(x) dx = 0$ and so $f = 0$.

2.24 Substitute $u = e^{-t}$ and $du = -e^{-t} dt$ which gives that $-\ln u = t$. Hence

$$\int_{0^+}^{1^-} (\ln u^{-1})^{-2^{-1}} du = \int_{0^+}^\infty e^{-t} t^{-2^{-1}} dt = \Gamma(2^{-1}) = \sqrt{\pi}.$$

2.25 If $n = 0$ then $\Gamma(2^{-1}) = \sqrt{\pi}$ as obtained above. Suppose that the formula is valid for $n = k$. When $n = k + 1$, we have

$$\Gamma(k+1+2^{-1}) = (k+2^{-1})\Gamma(k+2^{-1}) = (2(k+1))!\sqrt{\pi}(4^{k+1}(k+1)!)^{-1}.$$

as desired. The induction concludes the argument.

2.26 (i) This follows from easy calculations.

(ii) $L(af(x) + bg(x)) = a \int_{0+}^{\infty} e^{-xt} f(t) dt + b \int_{0+}^{\infty} e^{-xt} g(t) dt = aL(f(x)) + bL(g(x)).$

(iii) $L(e^{ax}f(x)) = \int_{0+}^{\infty} e^{-(x-a)t} f(t) dt = L(f(x-a)).$

(iv) Integration by parts yields

$$\int_{0+}^b e^{-xt} f'(t) dt = e^{-xb} f(b) - f(0) + \int_{0+}^b x e^{-xt} f(t) dt.$$

Since $\lim_{b \rightarrow \infty} e^{-xb} f(b) = 0$, one has

$$L(f'(x)) = \int_{0+}^{\infty} e^{-xt} f'(t) dt = -f(0) + \int_{0+}^{\infty} x e^{-xt} f(t) dt = -f(0) + xL(f(x)).$$

(v) This follows from that $L(x^n e^{ax}) = L((x-a)^n) = \frac{n!}{(x-a)^{n+1}}$ for $n \in \mathbb{N}$, thanks to $L(x^n) = \frac{n!}{x^{n+1}}$, $x > 0$.

3 Riemann-Stieltjes Integrals

3.1 Note that the total variation of a monotonic function is given by $V_x^y f = |f(y) - f(x)|$. So we can find $V_a^b f$ in general by partitioning $[a, b]$ into subintervals on which f is monotonic.

(i)

$$\begin{aligned} V_0^{3\pi} f &= |f(\pi/2) - f(0)| + |f(3\pi/2) - f(\pi/2)| \\ &+ |f(5\pi/2) - f(3\pi/2)| + |f(3\pi) - f(5\pi/2)| = 6. \end{aligned}$$

(ii) $f'(x) = 6x(x-1) = 0$ implies $x = 0$ and $x = 1$. Hence

$$V_{-1}^2 f = |f(0) - f(-1)| + |f(1) - f(0)| + |f(2) - f(1)| = 11.$$

(iii) If $1/(n+1) < \delta < 1/n$, then on $[\delta, 1]$ the only contributions to $V_{\delta}^1 f$ occur at $x = 1/(k+1)$, $k = 1, \dots, n-1$, where

$$\left| f\left(\frac{1}{k+1}\right) - f\left(\frac{1}{k+1} - \varepsilon_k\right) \right| + \left| f\left(\frac{1}{k+1} + \varepsilon_k\right) - f\left(\frac{1}{k+1}\right) \right| = 2|a_k|.$$

Here $\varepsilon_k > 0$ is such that

$$1/(k+2) < 1/(k+1) - \varepsilon_k < 1/(k+1) < 1/(k+1) + \varepsilon_k < k^{-1}.$$

Therefore $V_\delta^1 f = \sum_{k=1}^{n-1} 2|a_k|$. Letting $\delta \rightarrow 0$, we get $V_0^1 f = \sum_{k=1}^\infty 2|a_k|$.

3.2 First of all, let us prove that if f is monotonic on $[a, b]$ then $v_f(x) = |f(x) - f(a)|$ for $x \in [a, b]$. It suffices to prove this for increasing f . When P is any partition of $[a, x]$, one has

$$S(f, P) = \sum_{k=1}^n |f(x_k) - f(x_{k-1})| = \sum_{k=1}^n (f(x_k) - f(x_{k-1})) = f(x) - f(a).$$

Hence

$$v_f(x) = \sup\{S(f, P) : \text{partitions } P \text{ of } [a, x]\} = |f(x) - f(a)|.$$

Since $\sin x$ increases on $[0, \pi/2]$ and $[3\pi/2, 2\pi]$ and decreases on $[\pi/2, 3\pi/2]$, if $x \in [0, \pi/2]$ then $v_f(x) = \sin x$, if $x \in [\pi/2, 3\pi/2]$ then

$$v_f(x) = V_0^{\pi/2} f + V_{\pi/2}^x f = 2 - \sin x,$$

and if $x \in [3\pi/2, 2\pi]$ then

$$v_f(x) = V_0^{\pi/2} f + V_{\pi/2}^{3\pi/2} f + V_{3\pi/2}^x f = 4 + \sin x.$$

3.3 This follows from $L(P) = \sum_{k=1}^n ((x(t_k) - x(t_{k-1}))^2 + (y(t_k) - y(t_{k-1}))^2)^{\frac{1}{2}}$.

3.4 If $C(f_1)$ and $C(f_2)$ are the sets of all continuous points of f_1 and f_2 respectively, then for $x_0 \in C(f_1)$ we have $f_1(x_0^+) = f_1(x_0) = f_1(x_0^-)$. Since D is dense in a, b , there are points $x_j \in D$ such that $x_j \leq x_0$ and $\lim_{j \rightarrow \infty} x_j = x_0$, and consequently,

$$f_1(x_0^-) = \lim_{j \rightarrow \infty} f_1(x_j) = \lim_{j \rightarrow \infty} f_2(x_j) = f_2(x_0^-).$$

In a similar manner, we have $f_1(x_0^+) = f_2(x_0^+)$ and $x_0 \in C(f_2)$. Accordingly, $C(f_1) \subseteq C(f_2) \subseteq C(f_1)$.

3.5 Let $f = f_1$ and $g = f_2$. Within the k th term of the sum

$$S(gf, P) = \sum_{k=1}^n |g(x_k)f(x_k) - g(x_{k-1})f(x_{k-1})|$$

add $-g(x_k)f(x_{k-1}) + g(x_k)f(x_{k-1})$, and factor by pairs. This leads to

$$S(gf, P) \leq \sum_{k=1}^n |g(x_k)| |f(x_k) - f(x_{k-1})| + \sum_{k=1}^n |f(x_{k-1})| |g(x_k) - g(x_{k-1})|,$$

and so $S(gf, P) \leq M_g S(f, P) + M_f S(g, P)$, where $M_g = \sup_{x \in [a, b]} |g(x)|$ and $M_f = \sup_{x \in [a, b]} |f(x)|$.

3.6 Refer to Example 3.1.4

3.7 Let $P = \{x_k\}_{k=0}^n$ be any partition of $[a, b]$. Then

$$\begin{aligned} \sum_{k=1}^n |f(x_k) - f(x_{k-1})| &= \sum_{k=1}^n \lim_{j \rightarrow \infty} |f_j(x_k) - f_j(x_{k-1})| \\ &= \lim_{j \rightarrow \infty} \sum_{k=1}^n |f_j(x_k) - f_j(x_{k-1})| \\ &\leq \sup_{j \in \mathbf{N}} V_a^b(f_j) < \infty, \end{aligned}$$

and hence f is of bounded variation on $[a, b]$.

3.8 (i) Since $[x]$ is a step function, each sum reduces to those terms consisting of x^2 evaluated near the jump points times the jump of $[x]$ at such an x . Thus

$$\int_0^4 x^2 d([x]) = 1^2(1-0) + 2^2(2-1) + 3^2(3-2) + 4^2(4-3) = 30.$$

(ii) $\int_0^1 x^3 d(x^2) = 2 \int_0^1 x^4 dx = 2/5$.

(iii) $\int_a^c f dg = f(b)(g(b^+) - g(b)) = f(b)(1-0) = f(b)$.

3.9 It is enough to prove Theorem 3.2.4 (i).

$$\begin{aligned} &\int_a^b (f_1(x) + f_2(x)) dg_1(x) \\ &= \lim_{\|P\| \rightarrow 0} \sum_{k=1}^n (f_1(\xi_k) + f_2(\xi_k)) (g_1(\xi_k) - g_1(\xi_{k-1})) \\ &= \lim_{\|P\| \rightarrow 0} \left(\sum_{k=1}^n (f_1(\xi_k) (g_1(\xi_k) - g_1(\xi_{k-1}))) + \sum_{k=1}^n (f_2(\xi_k) (g_1(\xi_k) - g_1(\xi_{k-1}))) \right) \\ &= \int_a^b f_1 dg_1 + \int_a^b f_2 dg_1. \end{aligned}$$

3.10

$$\int_0^{\pi/2} x d(\sin x) = \pi/2 - \int_0^{\pi/2} \sin x dx = \pi/2 - 1.$$

3.11 If P is a partition of $[a, b]$ such that $c \notin P$, then the Riemann-Stieltjes sums are the same for $\int_a^b f dh$ as for $\int_a^b f dg$. Consider a partition P of $[a, b]$ such that $c = x_m \in P$. With the abbreviation $\Delta h_k = h(x_k) - h(x_{k-1})$, we have

$$\begin{aligned} \sum_{k=1}^n f(t_k) \Delta h_k - \sum_{k=1}^n f(t_k) \Delta g_k \\ &= f(t_m) \Delta h_m + f(t_{m+1}) \Delta h_{m+1} - f(t_m) \Delta g_m - f(t_{m+1}) \Delta g_{m+1} \\ &= f(t_m) (h(c) - g(c)) + f(t_{m+1}) (g(c) - h(c)) \\ &= (f(t_m) - f(t_{m+1})) (h(c) - g(c)). \end{aligned}$$

As $\|P\| \rightarrow 0$, t_m and t_{m+1} both tend to c , and the continuity of f yields a limit of 0. Hence, the Riemann-Stieltjes sums for $\int_a^b f dh$ approach the same limit as the Riemann-Stieltjes sums for $\int_a^b f dg$.

Note that this assertion fails if c is allowed to be either endpoint: e.g., on $[0, 1]$ take $f(x) = 1$, $g(x) = 0$ and

$$h(x) = \begin{cases} 0, & x \in [0, 1) \\ 1, & x = 1 \end{cases}.$$

This gives $\int_0^1 f(x) dg(x) = 0$ and $\int_0^1 f(x) dh(x) = 1$.

3.12 If f has bounded variation on $[a, b]$, then it is Riemann integrable there. The converse is false, however, which can be seen by function

$$f(x) = \begin{cases} x \sin \frac{\pi}{2x}, & x \neq 0 \\ 0, & x = 0 \end{cases}.$$

Clearly, this continuous (and therefore integrable) function f fails to have bounded variation.

3.13 Just choose $f = 1$ on $[0, 1]$ and

$$g(x) = \begin{cases} 1, & x \in (0, 1] \\ 0, & x = 0 \end{cases}.$$

3.14 (i) Let $P = \{x_k\}_{k=0}^n$ be a partition of $[a, b]$ and $\xi_k \in [x_{k-1}, x_k]$ for $k \in$

$\{1, 2, \dots, n\}$. Then

$$\begin{aligned} \left| \sum_{k=1}^n f(\xi_k) (g(x_k) - g(x_{k-1})) \right| &\leq \sum_{k=1}^n |f(\xi_k)| |g(x_k) - g(x_{k-1})| \\ &\leq \sum_{k=1}^n |f(\xi_k)| (v_g(x_k) - v_g(x_{k-1})) \\ &\leq \sup_{x \in [a, b]} |f(x)| \sum_{k=1}^n (v_g(x_k) - v_g(x_{k-1})), \end{aligned}$$

thereby giving the required inequalities.

(ii) and (iii) These two limits follow from (i) above.

3.15 (i) This follows from 3.14 (i).

(ii) For $x, x_0 \in [a, b]$, use

$$|F(x) - F(x_0)| = \left| \int_{x_0}^x f(t) dg(t) \right| \leq \sup_{t \in [a, b]} |f(t)| |g(x) - g(x_0)|.$$

(iii) For $x, x_0 \in [a, b]$ with $x \neq x_0$, using Theorem 3.4.3 (i) we get a point ξ between x_0 and x such that

$$F(x) - F(x_0) = \int_{x_0}^x f(t) dg(t) = f(\xi) (g(x) - g(x_0)),$$

and consequently,

$$F'(x_0) = \lim_{x \rightarrow x_0} f(\xi) g'(x_0) = f(x_0) g'(x_0)$$

provided that f is continuous at x_0 and g is differential at x_0 .

(iv) By (i) above it follows $h \in RS_F[a, b]$ because of $h \in C[a, b]$. To get that formula, let $P = \{x_k\}_{k=0}^n$ be a partition of $[a, b]$ and $\xi_k \in [x_{k-1}, x_k]$ for $k = 1, 2, \dots, n$. Then

$$\sum_{k=1}^n h(\xi_k) (F(x_k) - F(x_{k-1})) = \sum_{k=1}^n h(\xi_k) \int_{x_{k-1}}^{x_k} f(x) dg(x)$$

and

$$\int_a^b h(x) f(x) dg(x) = \sum_{k=1}^n \int_{x_{k-1}}^{x_k} h(x) f(x) dg(x).$$

Suppose that $M = 1 + \sup_{x \in [a, b]} |f(x)|$. Since h is uniformly continuous on $[a, b]$, for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$x, y \in [a, b] \text{ and } |x - y| < \delta \Rightarrow |h(x) - h(y)| < \eta = \frac{\varepsilon}{2M(1 + V_a^b g)}.$$

Accordingly, $\|P\| < \delta$, Problem 3.14 and Lemma 3.4 (i) give

$$\begin{aligned} & \left| \int_a^b h(x) f(x) dg(x) - \sum_{k=1}^n h(\xi_k) (F(x_k) - F(x_{k-1})) \right| \\ & \leq \sum_{k=1}^n \left| \int_{x_{k-1}}^{x_k} (h(x) - h(\xi_k)) f(x) dg(x) \right| \\ & \leq \sum_{k=1}^n \eta M V_{x_{k-1}}^{x_k} g = \eta M V_a^b g = 2^{-1} \varepsilon < \varepsilon, \end{aligned}$$

and the desired formula.

3.16 If $g \in C[0, 1]$ and $\int_0^1 f(x) dg(x) = 0$ for all increasing functions f on $[0, 1]$, then g must be a constant. In fact, first, $f = 1$ implies $g(0) = g(1)$; second, integrating by part, we have $\int_0^1 g(x) df(x) = (f(1) - f(0))g(0)$ and so $\int_0^1 (g(x) - g(0)) df(x) = 0$. If $f(x) = x^n$, $n \in \mathbb{N}$, then $\int_0^1 (g(x) - g(0)) x^{n-1} dx = 0$. Since $f \in C[0, 1]$, $g(x) - g(0) = 0$ follows from Problem 2.23.

3.17 Just integrate by part and use Remark 3.3.5.

3.18 (i) Modify the argument for Theorem 2.2.8 with $dg(x)$ replacing dx ;

(ii) Expand the left-hand integral;

(iii) Write the inside integral of the left-hand side of the formula in (ii) as $\int_{y \geq x} + \int_{y < x}(\cdots)$. Using the monotone condition on f_1 and f_2 , we find the integral is always nonnegative, reaching the required inequality;

(iv) This is similar to (iii).

3.19 (i) Use $x^{3/2} \leq x\sqrt{x+1} \leq \sqrt{2}x^{3/2}$ for $x \geq 1$ and $x\sqrt{x+1} \leq \sqrt{2}x$ for $0 \leq x \leq 1$;

(ii) Use $|\sin x^{-1}| \leq 1$ for $x > 0$;

(iii) Use the substitution $t = x^{-1}$ and Example 3.5.10.

4 Lebesgue-Radon-Stieltjes Integrals

4.1 $m_g((0, 1)) = 1 - e^{-1}$; $m_g([0, 1]) = 3 - e^{-1}$; $m_g((-1, 1)) = 4 - e^{-1}$; $m_g([0, 0]) = 1 - e^{-1}$.

4.2 (i) Let $S_1 = \cup_{k=1}^{n_1} I_{1,k}$ and $S_2 = \cup_{k=1}^{n_2} I_{2,k}$ be disjoint simple subsets of \mathbf{R} , where $\{I_{1,k}\}_{k=1}^{n_1}$ and $\{I_{2,k}\}_{k=1}^{n_2}$ are two sets of disjoint intervals. Then $S_1 \cup S_2 = \cup_{k=1}^{n_1+n_2} I_k$ where $I_k = I_{1,k}$ for $k = 1, 2, \dots, n_1$ and $I_{k+n_1} = I_{2,k}$ for $k = 1, 2, \dots, n_2$. Hence

$$m_g(S_1 \cup S_2) = \sum_{k=1}^{n_1+n_2} m_g(I_k) = m_g(S_1) + m_g(S_2).$$

But, let $g(x) = x$. Then $S_1 = [-1, 2^{-1}]$ and $S_2 = [-2^{-1}, 1]$ yield $S_1 \cap S_2 \neq \emptyset$ and

$$m_g(S_1 \cup S_2) = 2 \neq 3 = m_g(S_1) + m_g(S_2).$$

Meanwhile, $S_1 = [-1, 0]$ and $S_2 = [0, 1]$ give $S_1 \cap S_2 = \{0\}$ and

$$m_g(S_1 \cup S_2) = 2 = m_g(S_1) + m_g(S_2).$$

(ii) In this case $S_2 \setminus S_1$ is a simple subset of \mathbf{R} and $S_2 = S_1 \cup (S_2 \setminus S_1)$. Thus (i) is used to produce $m_g(S_2) = m_g(S_1) + m_g(S_2 \setminus S_1)$ and the required equality. But, under $g(x) = x$, $S_1 = (1, 3)$ and $S_2 = [2, 3] \Rightarrow m_g(S_2 \setminus S_1) = 0 \neq -1 = m_g(S_2) - m_g(S_1)$, and $S_1 = \{0\}$ and $S_2 = (0, 2] \Rightarrow m_g(S_2 \setminus S_1) = 2 = m_g(S_2) - m_g(S_1)$.

4.3 (i) Yes, $A_g(s_1) = 1$. (ii) Yes, $A_g(s_2) = 0$. (iii) Yes, $A_g(s_3) = 3$. (iv) Yes, $A_g(s_4) = -3$. (v) No.

4.4 Write $(0, 1) \cap \mathbf{Q}$ as $\{r_j\}_{j=1}^{\infty}$ and define

$$s_j(x) = \begin{cases} 0, & x \in [0, r_j) \\ 1, & x \in [r_j, r_j] \\ 0, & x \in (r_j, 1] \\ 0, & x \in \mathbf{R} \setminus [0, 1] \end{cases}.$$

If $f_k = \sum_{j=1}^k s_j$, then it is a step function on \mathbf{R} , but $f = \lim_{k \rightarrow \infty} f_k$ is not a step function on \mathbf{R} since it cannot be described via taking constant values on any finite set of subintervals of $[0, 1]$. Actually, if x is irrational in $(0, 1)$; equals 0 or 1; or belongs to $\mathbf{R} \setminus [0, 1]$, then $f_k(x) = 0 = f(x)$. If $x \in (0, 1)$ equals some r_{n_0} , then $f_k(r_{n_0}) = 1 = f(r_{n_0})$ when $k \geq n_0$. Note that $A_g(s_j) = r_j - r_j = 0$. So $S_g(f) = 0$.

4.5 (i)

$$\int_0^b f(x) dx = \begin{cases} \frac{b-[b]}{1+[b]}, & [b] \leq b < [b] + \frac{1}{2} \\ \frac{1+[b]-b}{1+[b]}, & [b] + \frac{1}{2} \leq b < [b] + 1 \end{cases}.$$

(ii) Use (i) to verify $|\int_0^b f(x) dx| \leq (2(1 + [b]))^{-1}$.

(iii) Note that if $[b] > 1$ then

$$\int_0^b |f(x)| dx = \sum_{k=1}^{[b]} k^{-1} + \frac{b - [b]}{1 + [b]}.$$

4.6 Use Theorem 4.2.3 (ii).

4.7 (i) Use Theorems 4.2.3 (ii) and 4.2.1.

(ii) Use $m_{g_1+g_2}(I) = m_{g_1}(I) + m_{g_2}(I)$ for any finite interval $I \subseteq \mathbf{R}$.

4.8 (i) Use $1_{E_1 \cup E_2} \leq 1_{E_1} + 1_{E_2}$ for $E_1, E_2 \subseteq \mathbf{R}$.

(ii) $E = \{0\}$ is not an m_g -null set where

$$g(x) = \begin{cases} 1, & x \in (-\infty, 0) \\ 0, & x \in [0, \infty) \end{cases}.$$

(iii) Use Theorem 4.2.3 (ii) and the definition of m_g -a.e. on \mathbf{R} .

4.9 (i) Use Theorem 4.3.1 and 4.2.3 (ii) for $\sum_{k=1}^n f_k$.

(ii) Use Theorem 4.3.2.

4.10 (i) $f = 0$.

(ii) $\lim_{n \rightarrow \infty} \int_{\mathbf{R}} f_n(x) dm_{id}(x) = 1 \neq 0 = \int_{\mathbf{R}} f(x) dm_{id}(x)$.

(iii) $\{f_n\}_{n=1}^{\infty}$ is neither increasing nor decreasing.

4.11 (i) Since $1 - \frac{x}{n} \leq e^{-\frac{x}{n}}$, the following sequence of functions

$$f_n(x) = \begin{cases} (1 - \frac{x}{n})^n, & x \in [0, n] \\ 0, & x \in (n, \infty) \cup (-\infty, 0) \end{cases}$$

is increasing and bounded from above by $1_{\mathbf{R}} e^{-x}$. The monotone convergence theorem ensures the desired limit formula.

(ii) By the dominated convergence theorem we have

$$\begin{aligned} \int_{\mathbf{R}} 1_{(0, \infty)}(x) |x|^\alpha (e^x - 1)^{-1} dm_{id}(x) &= \int_{\mathbf{R}} 1_{(0, \infty)}(x) |x|^\alpha \sum_{j=1}^{\infty} e^{-jx} dm_{id}(x) \\ &= \sum_{j=1}^{\infty} \int_{\mathbf{R}} 1_{(0, \infty)}(x) |x|^\alpha e^{-jx} dm_{id}(x) \\ &= \sum_{j=1}^{\infty} j^{-(1+\alpha)} \Gamma(1 + \alpha). \end{aligned}$$

4.12 (i) For each $n \in \mathbf{N}$ let

$$f_n(x) = \begin{cases} 0 & , \quad x \in (-\infty, n^{-1/p}) \\ x^{-p} & , \quad x \in [n^{-1/p}, 1] \end{cases}.$$

Then $\{f_n\}_{n=1}^\infty$ is increasing and $\lim_{n \rightarrow \infty} f_n(x) = f(x) = 1_{(0,1]}(x)|x|^{-p}$ for each $x \in \mathbf{R}$. By the monotone convergence theorem we have that $\int_{\mathbf{R}} 1_{(0,1]}(x)|x|^{-p} dm_{id}(x)$ equals

$$\lim_{n \rightarrow \infty} \int_{\mathbf{R}} 1_{(0,1]}(x) f_n(x) dm_{id}(x) = \begin{cases} (1-p)^{-1} & , \quad p \in (0, 1) \\ \infty & , \quad p \in [1, \infty) \end{cases}.$$

(ii) For each $n-1 \in \mathbf{N}$ we have the following two estimates:

Case 1: $x \in (0, 1)$ implies

$$x^{-n-1} \left(1 + \frac{x}{n}\right)^{-n} \leq x^{-n-1} \leq x^{-2-1}.$$

Case 2: $x \in [1, \infty)$ implies

$$x^{-n-1} \left(1 + \frac{x}{n}\right)^{-n} \leq \left(1 + x + \frac{n-1}{2n}x^2\right)^{-1} \leq 4x^{-2}.$$

Now, if

$$h(x) = \begin{cases} 0 & , \quad x \in (-\infty, 0] \\ x^{-2-1} & , \quad x \in (0, 1) \\ 4x^{-2} & , \quad x \in [1, \infty) \end{cases},$$

then

$$\int_{\mathbf{R}} 1_{(0,\infty)}(x) h(x) dm_{id}(x) = \int_0^1 x^{-2-1} dx + 4 \int_1^\infty x^{-2} dx = 6.$$

Using the dominated convergence theorem we obtain

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int_{\mathbf{R}} 1_{(0,\infty)}(x) |x|^{-n-1} \left(1 + \frac{|x|}{n}\right)^{-n} dm_{id}(x) \\ &= \int_{\mathbf{R}} 1_{(0,\infty)}(x) \lim_{n \rightarrow \infty} |x|^{-n-1} \left(1 + \frac{|x|}{n}\right)^{-n} dm_{id}(x) \\ &= \int_{\mathbf{R}} 1_{(0,\infty)}(x) e^{-x} dm_{id}(x) = 1. \end{aligned}$$

4.13 (i) If $x > 0$ then

$$f(x) = \sum_{n=1}^{\infty} e^{-nx} - 2 \sum_{n=1}^{\infty} e^{-2nx} = \frac{e^{-x} - e^{-2x}}{1 - e^{-2x}}.$$

(ii) Since $\int_0^\infty |f_n(x)|dx$ and $\int_0^\infty |f(x)|dx$ are convergent in the sense of improper Riemann integral, they are Lebesgue integrable on $(0, \infty)$.

(iii) $\int_{\mathbf{R}} 1_{(0, \infty)}(x) f(x) dm_{id}(x) \neq 0 = \sum_{n=1}^\infty \int_{\mathbf{R}} 1_{(0, \infty)}(x) f_n(x) dm_{id}(x)$.

4.14 (i) $N_1 = 1/4$; $N_2 = 0$; $N_3 = 3/4$; $N_4 = 1$.

(ii) We have $N_2 \leq N_1 \leq N_3 \leq N_4$ which is a version of Fatou's lemma.

4.15 Assume $\lim_{x \rightarrow \infty} f(x) \neq 0$. Then there is an $\epsilon_0 > 0$ such that for each $n \in \mathbf{N}$ there is an $x_n \in [n, \infty)$ obeying $|f(x_n)| \geq \epsilon_0$. Since f is uniformly continuous on $(0, \infty)$, there exists a $\delta > 0$ such that $|f(t_1) - f(t_2)| < 2^{-1}\epsilon_0$ whenever $t_1, t_2 \in (0, \infty)$ and $|t_1 - t_2| < \delta$. Accordingly, $|f(x) - f(x_n)| < 2^{-1}\epsilon_0$ whenever $x \in (x_n - \delta, x_n + \delta)$. Consequently, $|f(x)| \geq |f(x_n)| - 2^{-1}\epsilon_0 \geq 2^{-1}\epsilon_0$. Note that $x_n \geq n$. So there is a subsequence $(x_{n_k})_{k=1}^\infty$ of $\{x_n\}_{n=1}^\infty$ such that $x_{n_k} \leq x_{n_{k+1}}$ and $|x_{n_{k+1}} - x_{n_k}| > 2\delta$ for all $k \in \mathbf{N}$. If $I_k = (x_{n_k} - \delta, x_{n_k} + \delta)$, then $f \in LRS_g(0, \infty)$ implies

$$\infty > \int_{(0, \infty)} |f| dm_g \geq \int_{\cup_{k=1}^\infty I_k} |f| dm_g = \sum_{k=1}^\infty \int_{I_k} |f| dm_g \geq 2^{-1}\epsilon_0 \sum_{k=1}^\infty m_g(I_k) = \infty,$$

a contradiction.

4.16 (i) For any sequence $y_k \rightarrow y_0 \in [c, d]$, apply Theorem 4.3.1 to $f_k(\cdot) = f(\cdot, y_k)$.

(ii) Let $y_k \rightarrow y_0 \in [c, d]$, so for $x \in [a, b]$,

$$h_k(x) = \frac{f(x, y_k) - f(x, y_0)}{y_k - y_0} \rightarrow \frac{\partial f(x, y)}{\partial y}.$$

By the mean value theorem for derivative, we obtain

$$|h_k(x)| \leq \sup_{y \in [c, d]} \left| \frac{\partial f(x, y)}{\partial y} \right| \leq h(x)$$

thereby deriving the required formula via Theorem 4.3.1.

4.17 It is enough to consider the case $\sum_{j=1}^\infty m_g(E_j) < \infty$. Apply the monotone

convergence theorem with $f_n = \sum_{j=1}^n 1_{E_j}$ to obtain

$$\begin{aligned}
 m_g(\cup_{j=1}^{\infty} E_j) &= \int_{\mathbf{R}} 1_{\cup_{j=1}^{\infty} E_j} dm_g \\
 &\leq \int_{\mathbf{R}} \sum_{j=1}^{\infty} 1_{E_j} dm_g \\
 &= \lim_{n \rightarrow \infty} \int_{\mathbf{R}} \sum_{j=1}^n 1_{E_j} dm_g \\
 &= \sum_{j=1}^{\infty} m_g(E_j).
 \end{aligned}$$

4.18 If $\int_E |f| dm_g < \infty$ then

$$\begin{aligned}
 \infty &> \sum_{j=-\infty}^0 \int_{E_j} |f| dm_g + \sum_{j=1}^{\infty} \int_{E_j} |f| dm_g \\
 &\geq \sum_{j=-\infty}^0 |j| m_g(E_j) + \sum_{j=1}^{\infty} (j-1) m_g(E_j) \\
 &= \sum_{j=-\infty}^{\infty} |j| m_g(E_j) - \sum_{j=1}^{\infty} m_g(E_j) \\
 &= \sum_{j=-\infty}^{\infty} |j| m_g(E_j) - m_g(\cup_{j=1}^{\infty} E_j),
 \end{aligned}$$

and hence

$$\sum_{j=-\infty}^{\infty} |j| m_g(E_j) \leq \int_E |f| dm_g + m_g(E) < \infty.$$

Conversely, if $\sum_{j=-\infty}^{\infty} |j| m_g(E_j) < \infty$ then

$$\begin{aligned}
 \int_E |f| dm_g &= \sum_{j=-\infty}^0 \int_{E_j} |f| dm_g + \sum_{j=1}^{\infty} \int_{E_j} |f| dm_g \\
 &\leq \sum_{j=-\infty}^0 |j-1| m_g(E_j) + \sum_{j=1}^{\infty} j m_g(E_j) \\
 &\leq \sum_{j=-\infty}^{\infty} |j| m_g(E_j) + \sum_{j=-\infty}^0 m_g(E_j) \\
 &\leq \sum_{j=-\infty}^{\infty} |j| m_g(E_j) + m_g(E) < \infty.
 \end{aligned}$$

4.19 If f is m_g -measurable, then for any $c \in \mathbf{R}$,

$$\{x \in \mathbf{R} : f(x) \geq c\} = \bigcap_{j=1}^{\infty} \{x \in \mathbf{R} : f(x) > c - j^{-1}\}$$

is m_g -measurable.

Conversely, suppose $\{x \in \mathbf{R} : f(x) \geq c\}$ is m_g -measurable for any $c \in \mathbf{R}$. Then

$$\{x \in \mathbf{R} : f(x) > c\} = \bigcup_{j=1}^{\infty} \{x \in \mathbf{R} : f(x) \geq c + j^{-1}\}$$

is m_g -measurable and so is f .

4.20 (iii) For $n \in \mathbf{N}$ and $\varepsilon > 0$ let $E_{n,\varepsilon} = \{x \in \mathbf{R} : |f_n(x) - f(x)| \geq \varepsilon\}$. Then

$$\varepsilon m_g(\{x \in \mathbf{R} : |f_n(x) - f(x)| \geq \varepsilon\}) \leq \int_{E_{n,\varepsilon}} |f_n - f| dm_g \leq \int_{\mathbf{R}} |f_n - f| dm_g$$

gives the result.

(iv) For a Cauchy sequence $\{f_n\}_{n=1}^{\infty}$ in the m_g -measure, we choose a subsequence $\{f_{n_k}\}_{k=1}^{\infty}$ such that if $E_k = \{x \in \mathbf{R} : |f_{k+1}(x) - f_k(x)| \geq 2^{-k}\}$ then $m_g(E_k) \leq 2^{-k}$ and hence

$$m_g(\bigcup_{j=k}^{\infty} E_j) \leq \sum_{j=k}^{\infty} m_g(E_j) \leq \sum_{j=k}^{\infty} 2^{-j} = 2^{1-k}$$

and for $x \in \mathbf{R} \setminus \bigcup_{j=k}^{\infty} E_j$ and $p > q > k$,

$$|f_{n_p}(x) - f_{n_q}(x)| \leq \sum_{l=p}^{q-1} |f_{n_{l+1}}(x) - f_{n_l}(x)| \leq \sum_{l=p}^{q-1} 2^{-l} \leq 2^{1-p}.$$

The last estimate indicates that $\{f_{n_k}(x)\}_{k=1}^{\infty}$ is a Cauchy sequence for $x \in \mathbf{R} \setminus \bigcup_{j=k}^{\infty} E_j$. Now if

$$f(x) = \begin{cases} \lim_{k \rightarrow \infty} f_{n_k}(x), & x \in \mathbf{R} \setminus \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} E_j \\ 0, & x \in \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} E_j \end{cases},$$

then $f_{n_k} \rightarrow f$ m_g -a.e. since $m_g(\bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} E_j) = 0$; moreover $|f_{n_p}(x) - f(x)| \leq 2^{2-p}$ for $p \geq k$ and $x \notin \mathbf{R} \setminus \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} E_j$. Consequently, $f_{n_k} \rightarrow f$ in the m_g -measure and then $f_n \rightarrow f$ in the m_g -measure since

$$\begin{aligned} & \{x \in \mathbf{R} : |f_n(x) - f(x)| \geq \varepsilon\} \\ & \subseteq \{x \in \mathbf{R} : |f_n(x) - f_{n_k}(x)| \geq 2^{-1}\varepsilon\} \cup \{x \in \mathbf{R} : |f_{n_k}(x) - f(x)| \geq 2^{-1}\varepsilon\}. \end{aligned}$$

The uniqueness in the sense of m_g -a.e. follows from that if $f_n \rightarrow h$ in the m_g -measure then

$$\begin{aligned} & \{x \in \mathbf{R} : |h(x) - f(x)| \geq \varepsilon\} \\ & \subseteq \{x \in \mathbf{R} : |f_n(x) - h(x)| \geq 2^{-1}\varepsilon\} \cup \{x \in \mathbf{R} : |f_n(x) - f(x)| \geq 2^{-1}\varepsilon\} \end{aligned}$$

and hence $m_g(\{x \in \mathbf{R} : |h(x) - f(x)| \geq \varepsilon\}) = 0$ for any $\varepsilon > 0$.

(v) By Fatou's lemma we have

$$0 \leq \int_{\mathbf{R}} f dm_g \leq \liminf_{n \rightarrow \infty} \inf_{k \geq n} \int_{\mathbf{R}} f_k dm_g = 0$$

and thus $f = 0$ ϕ -a.e. on X thanks to $f \geq 0$ ϕ -a.e. on X .

(vi) If $f_n = 1_{(n, n+1]}$ and m_g is the Lebesgue measure on \mathbf{R} , then $f_n \rightarrow 0$ m_{id} -a.e. on \mathbf{R} , but f_n does not tend to 0 in the Lebesgue measure.

4.21 (i)

$$\int_{\mathbf{R}} f_1 dm_{id} = \int_{-1}^1 |x|^{-p} dx = 2 \int_0^1 x^{-p} dx = 2(1-p)^{-1}$$

and (via integration by part)

$$\int_{\mathbf{R}} f_2 dm_{id} = - \int_{-1}^1 \ln |x| dx = -2 \int_{0+}^1 \ln x dx = 2.$$

(ii)

$$\begin{aligned} \int_{\mathbf{R}} f_1 dm_{id} &= \int_0^\infty m_{id}(\{x \in \mathbf{R} : |x|^{-p} 1_{|x|<1} > t\}) dt \\ &= 2 \int_0^\infty \min\{1, t^{-1/p}\} dt \\ &= 2 \left(1 + \int_1^\infty t^{-1/p} dt \right) = 2/(1-p) \end{aligned}$$

and

$$\int_{\mathbf{R}} f_2 dm_{id} = \int_0^\infty m_{id}(\{x \in \mathbf{R} : (-\ln |x|) 1_{|x|<1} > t\}) dt = 2 \int_0^\infty e^{-t} dt = 2.$$

5 Absolute Continuities in Lebesgue Integrals

5.1 (i)&(ii) follow from considering $[a, b] \setminus E$ in Theorem 5.1.1. The equation of (iii) follows from (ii) and Theorem 5.1.1(iii). For the converse part, we use the fact that there are increasing open sets $\{O_j\}_{j=1}^\infty$ and decreasing closed sets $\{C_j\}_{j=1}^\infty$ such that $\cap_{j=1}^\infty C_j \supseteq E \supseteq \cup_{j=1}^\infty O_j$ to obtain

$$m_{id}(\cap_{j=1}^\infty (O_j \setminus E)) \leq m_{id}(\cap_{j=1}^\infty (O_j \setminus \cup_{k=1}^\infty O_k)) = 0.$$

Therefore

$$E = (\cap_{j=1}^\infty O_j) \setminus (\cap_{j=1}^\infty (O_j \setminus E))$$

is m_{id} -measurable.

5.2 If $E, F \subseteq \mathbf{R}$ are m_{id} -measurable, then $E \cup F = (E \setminus F) \cup (E \cap F) \cup (F \setminus E)$ and $E \cap F, F \setminus E, E \cap F$ are pairwise disjoint, and hence

$$m_{id}(E \cup F) = m_{id}(E \cap F) + m_{id}(F \setminus E) + m_{id}(E \setminus F).$$

In general, let $E, F \subseteq \mathbf{R}$ and $\varepsilon > 0$ be sufficiently small. Then according to Theorem 5.1.3(ii) there are two open sets O_e and O_f such that $O_e \supseteq E$ and $O_f \supseteq F$ such that

$$0 < m_{id}(O_e) - m_{id,*}(E) < \varepsilon \text{ and } 0 < m_{id}(O_f) - m_{id,*}(F) < \varepsilon.$$

Since

$$O_e \cap O_f \supseteq E \cap F \text{ and } O_e \cup O_f \supseteq E \cup F,$$

it follows that

$$\begin{aligned} m_{id,*}(E \cup F) + m_{id,*}(E \cap F) &\leq m_{id}(O_e \cup O_f) + m_{id}(O_e \cap O_f) \\ &= m_{id}(O_e) + m_{id}(O_f) \\ &\leq m_{id,*}(E) + m_{id,*}(F) + 2\varepsilon. \end{aligned}$$

5.3

$$\begin{aligned} m_{*,id}(E) &= \sup \{b - a - m_{id}([a, b] \setminus C) : \text{closed } C \subseteq E\} \\ &= b - a - \inf \{m_{id}([a, b] \setminus C) : \text{closed } C \subseteq E\} \\ &= b - a - \inf \{m_{id}(O) : \text{open } O \supseteq [a, b] \setminus E\} \\ &= b - a - m_{id,*}([a, b] \setminus E). \end{aligned}$$

5.4 This follows from Definition 5.1.2.

5.5 Let $\{a = x_0 < x_1 < \dots < x_n = x\}$ be a partition of the interval $[a, x]$. If $x < y$, then $\{a = x_0 < x_1 < \dots < x_n = x < x_{n+1} = y\}$ is a partition of the interval $[a, y]$, and hence

$$\sum_{j=1}^n (f(x_j) - f(x_{j-1}))^\pm \leq \sum_{j=1}^{n+1} (f(x_j) - f(x_{j-1}))^\pm \quad \text{and} \quad \langle V^\pm \rangle_a^x \leq \langle V^\pm \rangle_a^y.$$

Note that $u = u^+ - u^-$ holds for all $u \in \mathbf{R}$. So

$$\begin{cases} \sum_{j=1}^n (f(x_j) - f(x_{j-1}))^+ = \sum_{j=1}^n (f(x_j) - f(x_{j-1}))^- + (f(x) - f(a)); \\ \sum_{j=1}^n (f(x_j) - f(x_{j-1}))^- = \sum_{j=1}^n (f(x_j) - f(x_{j-1}))^+ + (f(a) - f(x)). \end{cases}$$

Accordingly,

$$\begin{cases} \langle V^+ \rangle_a^x f \leq \langle V^- \rangle_a^x f + (f(x) - f(a)); \\ \langle V^- \rangle_a^x f \leq \langle V^+ \rangle_a^x f + (f(a) - f(x)); \\ \text{(ii-3) holds.} \end{cases}$$

To verify ((ii-2)), just use $|u| = 2u^+ - u$ and ((ii-3)) to derive

$$\sum_{j=1}^n |f(x_j) - f(x_{j-1})| = 2 \sum_{j=1}^n (f(x_j) - f(x_{j-1}))^+ + (f(a) - f(x))$$

and

$$V_a^x f \geq 2\langle V^+ \rangle_a^b f + f(a) - f(x) = \langle V^+ \rangle_a^x f + \langle V^- \rangle_a^x f$$

whose case $x = b$ implies ((ii-2)).

For (iii), note that

$$\begin{cases} \int_a^b (\langle V^+ \rangle_a^x f)' dm_{id} \leq \langle V^+ \rangle_a^b f; \\ \int_a^b (\langle V^- \rangle_a^x f)' dm_{id} \leq \langle V^- \rangle_a^b f; \\ f'(x) = (\langle V^+ \rangle_a^x f)' - (\langle V^- \rangle_a^x f)'; \\ \text{(ii-2) holds.} \end{cases}$$

So, we get

$$f' \in LRS_{id}[a, b] \quad \text{with} \quad \int_a^b |f'| dm_{id} \leq V_a^b f.$$

Moreover, if $f \in AC[a, b]$ and $\{a = x_0 < x_1 < \dots < x_n = b\}$ is any partition of $[a, b]$, then Theorem 5.3.4 ensures $f' \in LRS_{id}[x_{j-1}, x_j]$ and

$$\begin{aligned} \sum_{j=1}^n |f(x_j) - f(x_{j-1})| &= \sum_{j=1}^n \left| \int_{x_{j-1}}^{x_j} f' dm_{id} \right| \\ &\leq \sum_{j=1}^n \int_{x_{j-1}}^{x_j} |f'| dm_{id} \\ &= \int_a^b |f'| dm_{id}, \end{aligned}$$

and hence

$$V_a^b f \leq \int_a^b |f'| dm_{id}$$

which implies the desired equality.

(iv) Since everywhere differentiable function f enjoys $f' \in LRS_{id}[a, b]$, for $\varepsilon > 0$ there is $\delta > 0$ such that $m_{id}^*(E) < \delta \Rightarrow \int_E |f'| dm_{id} < \varepsilon$. Now, if $\{[a_j, b_j]\}_{j=1}^n$ is a family of pairwise disjoint subintervals of $[a, b]$ with $\sum_{j=1}^n (b_j - a_j) < \delta$, then

$$\begin{aligned} \sum_{j=1}^n |f(b_j) - f(a_j)| &= \sum_{j=1}^n m_{id}(f([a_j, b_j])) \\ &\leq \sum_{j=1}^n \int_{a_j}^{b_j} |f'| dm_{id} \\ &= \int_{\cup_{j=1}^n [a_j, b_j]} |f'| dm_{id} < \varepsilon, \end{aligned}$$

and hence $f \in AC[a, b]$.

5.7 If $E \subseteq [a, b]$ is m_{id} -measurable, then there is a sequence $\{C_j\}$ of closed sets containing E with $m_{id}(E) = \lim_{j \rightarrow \infty} m_{id}(C_j)$, and hence there is an m_{id} -null set Z such that $E = Z \cup (\cup_{j=1}^{\infty} C_j)$. This, plus $f \in AC[a, b]$, implies

$$m_{id}(f(Z)) = 0 \text{ and } f(E) = f(Z) \cup f(\cup_{j=1}^{\infty} C_j) = f(Z) \cup (\cup_{j=1}^{\infty} f(C_j)).$$

Since $f \in AC[a, b]$ and C_j is compact, $f(C_j)$ is compact and consequently m_{id} -measurable. Accordingly, $f(E)$ is m_{id} -measurable.

5.7 Define $g(x) = \int_a^x f' dm_{id}$ and $h = f - g$. Since f is increasing on $[a, b]$, $g \in AC[a, b]$ and $h' = f' - g' = 0$ m_{id} -a.e. on $[a, b]$. If \tilde{g}, \tilde{h} obey $f = g + h = \tilde{g} + \tilde{h}$,

then $(g - \tilde{g})' = (\tilde{h} - h)' = 0$, and hence $\tilde{g} = g + \kappa$ and $\tilde{h} = h + \tau$ for a constant pair $\{\kappa, \tau\}$.

5.8 Under $f \in LRS_{id}[a, b]$, the if-part is trivial, but for the only-if part we can differentiate $\int_a^x f dm_{id} = 0$ to get $f(x) = 0$ for m_{id} -a.e. $x \in [a, b]$.

5.9 It follows from a modification of the argument for Theorem 5.5.1(ii).

5.10 (i) follows from a direct computation. (ii) follows from the following calculation with $0 < x < 2^{-1}$:

$$Mf_2(x) \geq (2x)^{-1} \int_0^{2x} |f_2(y)| dy \geq (2x)^{-1} \int_0^x (y \ln^2 y)^{-1} dy \geq (2x \ln x)^{-1}.$$

To reach (iii). we apply the Hardy-Littlewood maximal theorem and the Fubini theorem to get

$$\begin{aligned} \int_{\mathbf{R}} (Mf)^p dm_{id} &= p \int_0^\infty m_{id}(\{x \in \mathbf{R} : Mf(x) > t\}) t^{p-1} dt \\ &\leq (4p) \int_0^\infty \left(\int_{\{x \in \mathbf{R} : |f(x)| > 2^{-1}t\}} |f| dm_{id} \right) t^{p-2} dt \\ &\leq (4p) \int_{\mathbf{R}} \left(\int_0^{2|f|} t^{p-2} dt \right) |f| dm_{id} \\ &= p(p-1)^{-1} 2^{p+1} \int_{\mathbf{R}} |f|^p dm_{id}. \end{aligned}$$

6 Metric Spaces

6.1 (i) Properties of Definition 6.1.1: (i), (ii), and (iii) hold.

(ii) Properties of Definition 6.1.1: (i), (ii), and (iii) hold.

(iii) Properties of Definition 6.1.1: (i) and (iii) hold, but (ii) does not hold since taking $x = (1, 2)$ and $y = (0, 4)$ yields $d(x, y) = 3 \neq \sqrt{5} = d(y, x)$. An effort on verifying (iii) is strongly needed. Below is the details: Given $x = (x_1, x_2), y = (y_1, y_2), z = (z_1, z_2)$. Let us consider the following cases.

Case 1. $x_2 \geq z_2$: This implies $d(x, z) = ((x_1 - z_1)^2 + (x_2 - z_2)^2)^{2^{-1}}$.

Subcase 1.1. $y_2 \geq z_2$: This gives $d(y, z) = ((y_1 - z_1)^2 + (y_2 - z_2)^2)^{2^{-1}}$. And $d(x, y)$ equals $((x_1 - y_1)^2 + (x_2 - y_2)^2)^{2^{-1}}$ or $|x_1 - y_1| + |x_2 - y_2|$. Then by the triangle inequality for the Euclidean space (\mathbf{R}^2, d_2) and

$$|x_1 - y_1| + |x_2 - y_2| \geq ((x_1 - y_1)^2 + (x_2 - y_2)^2)^{2^{-1}}$$

one has $d(x, y) + d(y, z) \geq d(x, z)$.

Subcase 1.2. $y_2 < z_2$: This yields $d(y, z) = |y_1 - z_1| + |y_2 - z_2|$. And $d(x, y)$ equals $((x_1 - y_1)^2 + (x_2 - y_2)^2)^{2^{-1}}$ or $|x_1 - y_1| + |x_2 - y_2|$. Then by the triangle inequalities for both the Euclidean space distance d_2 and that one defined by Problem 6.1 (i), as well as

$$\begin{cases} |y_1 - z_1| + |y_2 - z_2| \geq ((y_1 - z_1)^2 + (y_2 - z_2)^2)^{2^{-1}}; \\ |x_1 - z_1| + |x_2 - z_2| \geq ((x_1 - z_1)^2 + (x_2 - z_2)^2)^{2^{-1}}, \end{cases}$$

one has $d(x, y) + d(y, z) \geq d(x, z)$.

Case 2. $x_2 < z_2$: This infers $d(x, z) = |x_1 - z_1| + |x_2 - z_2|$.

Subcase 2.1. $y_2 \geq z_2$: This means $d(y, z) = ((y_1 - z_1)^2 + (y_2 - z_2)^2)^{2^{-1}}$. Note that $x_2 < z_2$ and $z_2 \leq y_2$ imply $x_2 < y_2$ and $|x_2 - y_2| > |z_2 - y_2|$. So $d(x, y) = |x_1 - y_1| + |x_2 - y_2|$. Hence by the triangle inequality of the distance for \mathbf{R} ,

$$\begin{aligned} d(x, y) + d(y, z) &\geq |x_1 - y_1| + |x_2 - y_2| + |y_1 - z_1| \\ &\geq |x_1 - z_1| + |x_2 - z_2| = d(x, z). \end{aligned}$$

Subcase 2.2. $y_2 < z_2$: This tells us that $d(y, z) = |y_1 - z_1| + |y_2 - z_2|$ and thus $d(x, y)$ equals $((x_1 - y_1)^2 + (x_2 - y_2)^2)^{2^{-1}}$ or $|x_1 - y_1| + |x_2 - y_2|$. For the second case, one applies the triangle inequality of the distance for \mathbf{R} to get

$$d(x, y) + d(y, z) \geq |x_1 - z_1| + |x_1 - z_2| = d(x, z).$$

Concerning the first case: $x_2 \geq y_2$, note that $d(x, y) + d(y, z) \geq d(x, z)$ is equivalent to

$$((x_1 - y_1)^2 + (x_2 - y_2)^2)^{2^{-1}} + |y_1 - z_1| + |z_2 - y_2| \geq |x_1 - z_1| + |z_2 - x_2|$$

which is

$$|y_1 - z_1| + |x_2 - y_2| \geq |x_1 - z_1| - ((x_1 - y_1)^2 + (x_2 - y_2)^2)^{2^{-1}}.$$

Since

$$((x_1 - y_1)^2 + (x_2 - y_2)^2)^{2^{-1}} \geq |x_1 - y_1|,$$

one has

$$|x_1 - z_1| - ((x_1 - y_1)^2 + (x_2 - y_2)^2)^{2^{-1}} \leq |x_1 - z_1| - |x_1 - y_1| \leq |y_1 - z_1|.$$

Note that $x_2 \geq y_2$, so the desired estimate follows.

6.2 Suppose that S is the set of all positive integers $m \geq 2$ such that

$$d(x^{(1)}, x^{(m)}) \leq \sum_{j=2}^m d(x^{(j-1)}, x^{(j)})$$

for any set of m points in X . For $m = 2$ the result is trivial, and for $m = 3$ the assertion is the triangle inequality for (X, d) . Assume $m \in S$ and let $\{x^{(j)}\}_{j=1}^{m+1}$ be any $m+1$ points in X . By the triangle inequality for (X, d) , and the above assumption, one has

$$d(x^{(1)}, x^{(m+1)}) \leq d(x^{(1)}, x^{(m)}) + d(x^{(m)}, x^{(m+1)}) \leq \sum_{j=2}^{m+1} d(x^{(j-1)}, x^{(j)}).$$

Therefore $m+1 \in S$, and the mathematical induction ensures that $S = \mathbf{N} \setminus \{1\}$.

6.3 (i) It is enough to check the second result. Let $z^{(0)} = (x_1, x_2, \dots, x_n) = x$, $z^{(1)} = (y_1, x_2, \dots, x_n)$, $z^{(2)} = (y_1, y_2, \dots, x_n), \dots$, and $z^{(n)} = (y_1, y_2, \dots, y_n) = y$. Then for $j = 1, 2, \dots, n$, one has

$$d_n^*(z^{(j-1)}, z^{(j)}) = |x_j - y_j| = d_n(z^{(j-1)}, z^{(j)}),$$

and by adding these equations, one gets

$$d_n^*(z^{(0)}, z^{(n)}) = \sum_{j=1}^n d_n^*(z^{(j-1)}, z^{(j)}) \geq d_n(z^{(0)}, z^{(n)}).$$

(ii) This is a routine argument. Clearly, $d_{*,n}(x, y) \leq d_n(x, y) \leq \sqrt{n}d_{*,n}(x, y)$.

6.4 First, let's prove that A° is the largest open subset of A . If $O \subseteq A$ is open, then any point of O is an interior point of A , and so this point belongs to A° . This gives $O \subseteq A^\circ$. Second, let $K \supseteq A$ be closed. If $x \in \bar{A}$ then it suffices to prove that $x \in K$ when x is a limit point of A . In this case, x is a limit point of K . But since K is closed, $x \in K$. We are done.

6.5 Let $x \in \bar{B}_1(0)$ and $\varepsilon > 0$. We must show that $B_\varepsilon(x) \cap B_1(0) \neq \emptyset$. Choose a number c such that $1 - \varepsilon < c < 1$, and consider the point $z = cx = (cx_1, cx_2)$; this gives $d_2(z, 0) = cd_2(x, 0) < 1$, so $x \in B_1(0)$. Also, $\underline{d_2}(x, z) = (1 - c)d_2(x, 0) < \varepsilon$, so $x \in B_\varepsilon(x)$. Hence, every point of $\bar{B}_1(0)$ is in $\underline{B_1(0)}$, the closure of $B_1(0)$. Conversely, if $y \notin \bar{B}_1(0)$, then $d_2(y, 0) > 1$, and if $\delta = d_2(y, 0) - 1$, then by the

triangle inequality, $B_\delta(y) \cap B_1(0) = \emptyset$. Hence, only the points of $\overline{B_1}(0)$ are in $\overline{B_1}(0)$.

6.6 Let $p \in S$ and define $S_1 = \{p\}$ and $S_2 = S \setminus \{p\}$. Since finite sets have no limit points, it follows that the definition of disconnectness is satisfied.

6.7 This follows from the facts that $\overline{G_1} \cap G_2 = \emptyset$ and $\overline{G_2} \cap G_1 = \emptyset$.

6.8 (i) The sequence is convergent to $(1, 1)$;

(ii) The sequence is convergent to $(0, 0)$;

(iii) The sequence is divergent since $k \rightarrow \infty$.

6.9 For $k \in \mathbf{N}$ take $x_k = (2k)^{-1}$. Since $x_k \rightarrow 0$ but $0 \notin (0, 1)$.

6.10 (i) If $A = \{x^{(k)}\}_{k=1}^\infty$ is bounded then there is an $M > 0$ such that $d_n(x, 0) \leq M$ for all $x = (x_1, \dots, x_n) \in A$ and hence $\sup_{1 \leq j \leq n} \sup_{k \in \mathbf{N}} |x_j^{(k)}| \leq M$. The converse is trivial.

(ii) This follows easily from the inclusion between cubes and balls in \mathbf{R}^n .

6.11 Assume that $\{x^{(k)}\}_{k=1}^\infty$ is a Cauchy sequence and apply the definition of Cauchy sequence with $\epsilon \in (0, 1)$. For this distance function, $d(x, y) < 1$ implies $d(x, y) = 0$ and hence $x = y$, so we have that $k > m > n_0$ implies $x^{(k)} = x^{(m)}$. Hence, $k > n_0$ implies $x^{(k)} = x^{(n_0+1)}$.

6.12 Let $B_r(p)$ be an arbitrary open ball in \mathbf{R}^n . Then $C_{r/\sqrt{n}}(p)$, the cube with side-length r/\sqrt{n} and center p , is contained in $B_r(p)$. Since the irrational numbers $\mathbf{R} \setminus \mathbf{Q}$ are dense in \mathbf{R} , for each $j = 1, \dots, n$, the interval $(p_j - r/\sqrt{n}, p_j + r/\sqrt{n})$ contains an $x_j \in \mathbf{R} \setminus \mathbf{Q}$. Thus $|x_j - p_j| < r/\sqrt{n}$, which defines a point $x \in C_{r/\sqrt{n}}(p)$.

6.13 This follows from the triangle inequality for d .

6.14 Let $\{x_k\}_{k=1}^\infty$ be a Cauchy sequence in X . Then there is $n_1 \in \mathbf{N}$ such that $d(x_n, x_{n_1}) < 2^{-1}$ as $n \geq n_1$. Suppose x_{n_1}, \dots, x_{n_k} have been chosen such that $n_{j+1} > n_j$, $j = 1, \dots, k-1$ and $d(x_n, x_{n_j}) < 2^{-j}$ as $n \geq n_j$. Take $n_{k+1} > n_k$ so that $d(x_n, x_{n_{k+1}}) < 2^{-(k+1)}$ for $n \geq n_{k+1}$. Consider $\overline{B_{2^{-(k-1)}}}(x_{n_k})$. We claim that this sequence of closed balls is nested. If $x \in \overline{B_{2^{-k}}}(x_{n_{k+1}})$, then $d(x, x_{n_{k+1}}) \leq 2^{-k}$ and $d(x, x_{n_k}) \leq d(x, x_{n_{k+1}}) + d(x_{n_{k+1}}, x_{n_k}) < 2^{-(k-1)}$. So $x \in \overline{B_{2^{-(k-1)}}}(x_{n_k})$. By hypothesis there is $x \in \bigcap_{k=1}^\infty \overline{B_{2^{-(k-1)}}}(x_{n_k})$. Obviously, $x_{n_k} \rightarrow x$ and so $x_n \rightarrow x$. That is, X is complete.

Conversely, suppose that X is complete. If $\overline{B_{r_k}}(x_k)$ with $r_k \rightarrow 0$ is nested, then $\{x_k\}_{k=1}^\infty$ is a Cauchy sequence in X . In fact, $k > m$ implies $\overline{B_{r_k}}(x_k) \subseteq$

$\overline{B}_{r_m}(x_m)$, and so $x_k \in \overline{B}_{r_m}(x_m)$, i.e., $d(x_k, x_m) \leq r_m$. For any $\varepsilon > 0$ there is an $n_0 \in \mathbf{N}$ such that $m > n_0$ implies $r_m < \varepsilon$. Consequently, $k > m > n_0$ yields $d(x_k, x_m) < \varepsilon$. Accordingly, $\{x_k\}$ is convergent to a point $x \in X$. But, each $\overline{B}_{r_k}(x_k)$ contains all but possibly a finite number of the $\{x_k : k \in \mathbf{N}\}$ and since $\overline{B}_{r_k}(x_k)$ is closed, it follows that $x \in \overline{B}_{r_k}(x_k)$ and so that x is in all $\overline{B}_{r_k}(x_k)$.

6.15 Since S_j 's are nonempty, there is a sequence $\{x_j\}_{j=1}^\infty$ such that $x_j \in S_j$ for any $j \in \mathbf{N}$. Given $\varepsilon > 0$, there exists an $N \in \mathbf{N}$ depending only on ε such that $\text{diam}(S_j) < \varepsilon$ for $j > N$. Thus, $d(x_m, x_n) < \varepsilon$ whenever $m, n \geq N$. This means that $\{x_j\}_{j=1}^\infty$ is a Cauchy sequence in the complete metric space (X, d) , and so that it converges, to x , say. Since $S_{j+1} \subseteq S_j$ and S_j is closed, the fact $x_j \in S_j$ must give $x \in \bigcap_{j=1}^\infty S_j$.

6.16 If S is nowhere dense then \overline{S} contains no sphere, i.e., $\overline{S}^\circ = \emptyset$. Suppose that O is any open ball in X . Then \overline{S}° does not contain O , and hence the open $O \setminus \overline{S}$ is not \emptyset . Consequently, there is an open ball O' with $O' \subseteq O \setminus \overline{S} \subseteq O$. This implies $O' \cap S = \emptyset$ since $S \subset \overline{S}$.

Conversely, assume that every open ball O contains an open ball O' such that $O' \cap S = \emptyset$. If $\overline{S}^\circ \neq \emptyset$, then there would be a point x in the open set \overline{S}° and hence an open ball $B_r(x) \subseteq \overline{S}$ centered at x with a radius $r > 0$. As a by-product, any point p in $B_r(x)$ is a limit point of S , and then any open ball centered at p and contained in $B_r(x)$ has a non-void intersection with S . This is against the above assumption. Therefore $\overline{S}^\circ = \emptyset$.

6.17 Suppose that $\{O_\beta\}$ is an open covering of $\bigcap_{\alpha \in I} E_\alpha$. Since E_α is compact, it is closed and $X \setminus \bigcap_{\alpha \in I} E_\alpha$ is open. Now $E_\alpha \subseteq (X \setminus \bigcap_{\alpha \in I} E_\alpha) \cup_\beta O_\beta$. By the compactness of E_α , there is a finite number n of O_β such that $E_\alpha \subseteq (X \setminus \bigcap_{\alpha \in I} E_\alpha) \cup_{j=1}^n O_{\beta_j}$, and consequently, $\bigcap_{\alpha \in I} E_\alpha \subseteq \bigcup_{j=1}^n O_{\beta_j}$. Therefore $\bigcap_{\alpha \in I} E_\alpha$ is compact.

6.18 Assume O is an open cover of $\{x_j\}_{j=0}^\infty$. Then there exists an open set $O \in O$ such that $x_0 \in O$. Hence $\lim_{n \rightarrow \infty} d(x_n, x_0) = 0$ indicates that there is an $n_0 \in \mathbf{N}$ such that $x_n \in O$ as $n > n_0$. For each $n \in \{1, \dots, n_0\}$ there exists an open set $O_n \in O$ such that $x_n \in O_n$. Now $\{O, O_1, \dots, O_{n_0}\}$ forms a subcover of $\{x_j\}_{j=0}^\infty$. So, $\{x_j\}_{j=0}^\infty$ is compact.

7 Continuous Mappings

7.1 Given $p = (p_1, p_2, p_3) \in \mathbf{R}^3$ and $\varepsilon > 0$, choose $\delta = \varepsilon$. Now $x = (x_1, x_2, x_3) \in \mathbf{R}^3$ and $d_3(x, p) < \delta$ imply

$$d_3(f(x), f(p)) = ((x_1 - p_1)^2 + (x_2 - p_2)^2)^{2^{-1}} \leq d_3(x, p) < \varepsilon.$$

7.2 The natural choice is $f(0) = 0$ due to $d_2(0, 0) = 0$. We must now prove that f is continuous at 0. Given $\varepsilon > 0$, choose $\delta = \varepsilon$. Now $d_2(x, 0) < \delta$ implies $|f(x) - f(0)| = d_2(x, 0) < \varepsilon$.

7.3 The range of f is $[0, \infty)$, so no point is mapped into $(-3, -2)$; hence $f^{-1}(S) = \emptyset$.

7.4 Since $0 \notin D$, $f(0) = 0$. For a fixed $x = (x_1, x_2, x_3) \in \mathbf{R}^3$, consider the points on the line segment $\{tx = (tx_1, tx_2, tx_3) \in \mathbf{R}^3 : t \in [0, 1]\}$. We want to find $\lim_{t \rightarrow 0^+} f(tx)$. If $x_3 \leq 0$, then $tx \notin D$ and $f(tx) = 0$ for all $t \in [0, 1]$. If $x_3 > 0$, then we claim that for t sufficiently small, $tx \notin D$ and therefore $f(tx) = 0$. This is seen by checking the coordinate inequality for $tx \in D$: $tx \in D$ if and only if $tx_3 < (tx_1)^2 + (tx_2)^2$, or $x_3 < t(x_1^2 + x_2^2)$. For any $x_3 > 0$, t can be chosen small enough to deny this. Therefore in either case, $f(tx) = 0$ for t sufficiently small. The discontinuity of f at 0 is shown by finding points of D that are in $B_\delta(0)$ for any $\delta > 0$.

7.5 $T((1, 1)) = (3, -1)$ yields $a_{11} + a_{12} = 3$ and $a_{21} + a_{22} = -1$; $T((1, -1)) = (1, 7)$ yields $a_{11} - a_{12} = 1$ and $a_{21} - a_{22} = 7$. Solving this system of equations, we get $a_{11} = 2$, $a_{12} = 1$, $a_{21} = 3$ and $a_{22} = -4$.

7.6 Since f is continuous on compact subset K , f assumes its minimum, i.e., $\min_{x \in K} f(x) = f(m)$ for $m \in K$. Taking $\delta = f(m) > 0$ gives the desired result.

7.7 Let $y \in [0, 1]$. If y is rational or irrational, then taking $x = y$ or $x = 1 - y$ gives $f(x) = y$. So f is onto. But f is continuous only at $x = 2^{-1}$. In fact, $|f(x) - f(2^{-1})| = |f(x) - 2^{-1}| = |x - 2^{-1}|$ gives the continuity. Let $x \in [0, 1] \setminus \{2^{-1}\}$. If x is rational then $|f(x_k) - f(x)| = |f(x_k) - x|$. When x_k is rational resp. irrational, one then has $|f(x_k) - x| = |x_k - x|$ resp. $|1 - x_k - x|$. Thus, once choosing that x_k is rational but x_{k+1} irrational, one cannot get $|f(x_k) - x| \rightarrow 0$ as $x_k \rightarrow x$.

7.8 (i) Taking $f_1 = 1$ and $f_2 = 0$ we have $d_{C[0,1]}(T(f_1), T(f_2)) = 1 = d_{C[0,1]}(f_1, f_2)$;

(ii) If $T(f) = f$, then $\int_0^x f(t) dt = f(x)$ and hence $f = 0$, giving the uniqueness;
 (iii) Since

$$T \circ T(f(x)) = \int_0^x \left(\int_0^y f(t) dt \right) dy = \int_0^x (x-t)f(t) dt,$$

it follows that $d_{C[0,1]}(T \circ T(f_1), T \circ T(f_2)) \leq 2^{-1} d_{C[0,1]}(f_1, f_2)$.

7.9 The result follows from $|f(b) - f(a)| = |f'(c)(b-a)| \leq \alpha|b-a|$ for some $c \in [a, b]$.

7.10 Since the series is convergent, there is an $n_0 \in \mathbf{N}$ such that

$$\alpha_{n_0} = \sup_{x_1 \neq x_2} d_X(T_{(n_0)}(x_1), T_{(n_0)}(x_2)) / d_X(x_1, x_2) < 1.$$

This means that $T_{(n_0)}$ is a contraction, and consequently, it has a unique fixed point thanks to Corollary 7.4.5, and so does T .

7.11 If $f(x) = d_X(x, T(x))$, then $f : K \rightarrow \mathbf{R}$ is continuous. Since K is compact and nonempty, there is a point $x_0 \in K$ such that $f(x_0) = \min_{x \in K} f(x)$. Consequently, $T(x_0) = x_0$ – otherwise we would have $f(T(x_0)) < f(x_0)$ which shows that $f(x_0)$ is not the minimum of f over K , a contradiction. The uniqueness is trivial.

7.12 Consider the map $T : C[0, 1] \rightarrow C[0, 1]$ given by

$$T(f) = h(x) + \int_0^x f(x-t) \exp(-t^2) dt.$$

For $f_1, f_2 \in C[0, 1]$, we have $d_{C[0,1]}(T(f_1), T(f_2)) \leq 2^{-1} \sqrt{\pi} d_{C[0,1]}(f_1, f_2)$. Since $C[0, 1]$ is complete under $d_{C[0,1]}(f_1, f_2)$, T is a contraction and hence there exists an $f \in C[0, 1]$ such that $T(f) = f$.

7.13 Show that the inverse f^{-1} of f exists and apply the Banach contraction theorem to f^{-1} .

7.14 The equivalences can be readily checked via the related definitions.

7.15 Check that f is not injective.

7.16 Outline of the proof: For (X, d_X) define two Cauchy sequences $\{x_k^{(1)}\}_{k=1}^\infty$ and $\{x_k^{(2)}\}_{k=1}^\infty$ in X to be equivalent provided $\lim_{k \rightarrow \infty} d_X(x_k^{(1)}, x_k^{(2)}) = 0$. Let Y be the space of all equivalence classes of Cauchy sequences in X with metric

$$d_Y(\{x_k^{(1)}\}_{k=1}^\infty, \{x_k^{(2)}\}_{k=1}^\infty) = \lim_{k \rightarrow \infty} d_X(x_k^{(1)}, x_k^{(2)}).$$

Then (Y, d_Y) is complete. Define a mapping f from X into Y by $f(x) = \{x_k\}_{k=1}^\infty \in Y$ where $x_k = x \in X$ for each $k \in \mathbf{N}$. Then f is an isometry and moreover $f(X)$ is dense in Y .

7.17 Use the triangle inequality: $d_X(x, y) \leq d_X(x_0, y) + d_X(x, x_0)$.

8 Normed Linear Spaces

8.1 Let $\Phi(t) = t^{\frac{1}{p}}$ for $t \geq 0$. Since $p^{-1} \in (0, 1)$, $\Phi''(s) < 0$ for all $s > 0$ and Φ is concave. Hence $\Phi(t) \leq \Phi(1) + \Phi'(1)(t - 1)$ or $t^{\frac{1}{p}} \leq 1 + \frac{t-1}{p} = \frac{t}{p} + \frac{1}{q}$. Setting $t = u^p v^{-q}$, where $u \geq 0$ and $v > 0$, we find $uv \leq \frac{u^p}{p} + \frac{v^q}{q}$ where $1 - q = -\frac{q}{p}$. Obviously, this inequality also holds when $v = 0$. The last inequality implies the Hölder inequality which, plus $|a + b|^p \leq |a + b|^{p-1}|a| + |a + b|^{p-1}|b|$ (for $a, b \in \mathbf{R}$), yields the Minkowski inequality.

8.2 (i) $n - 1$, since the following $n - 1$ vectors

$$(1, -1, 0, \dots, 0), (0, 1, -1, 0, \dots, 0), \dots, (0, 0, 0, \dots, 1, -1)$$

are linearly independent;

(ii) ∞ , since $1, t, t^2, \dots, t^n$ are linearly independent for any $n \in \mathbf{N}$;

(iii) ∞ , since $1, t, t^2, \dots, t^n$ are linearly independent for any $n \in \mathbf{N}$.

8.3 (i) Obviously, $\|f\|_p \leq \|f\|_\infty$ when $p \in [1, \infty)$ and $f \in C[0, 1]$, but there is no such a constant $\kappa > 0$ that $\|f\|_\infty \leq \kappa \|f\|_p$ for all $f \in C[0, 1]$. For $n - 2 \in \mathbf{N}$ let

$$f_n(t) = \begin{cases} n^{\frac{2}{p}} t^{\frac{1}{p}}, & 0 \leq t \leq \frac{1}{n} \\ n^{\frac{2}{p}} \left(\frac{2}{n} - t\right)^{\frac{1}{p}}, & \frac{1}{n} \leq t \leq \frac{2}{n} \\ 0 & , \frac{2}{n} \leq t \leq 1 \end{cases}.$$

It is easy to see that $\|f_n\|_\infty = n^{\frac{1}{p}} \rightarrow \infty$ (as $n \rightarrow \infty$) and $\|f_n\|_p = 1$. This verifies the nonexistence of the above constant $\kappa > 0$.

(ii) For $n \in \mathbf{N}$ let $f_n(x) = x^n$. Then $\|f_n\|_\infty = 1$ and hence the unit sphere of $(C[0, 1], \|\cdot\|_\infty)$ is not finite dimensional. Theorem 8.2.5 (ii) gives that the unit sphere is not compact.

8.4 Note that $\|x\| - \|a\| \leq \|x - a\|$. So $\|\cdot\| : X \rightarrow \mathbf{R}$ is continuous. Also, regarding the continuity of the vector addition and the scalar multiplication, we naturally assume that the norms defined on $X \times X$ and $\mathbf{F} \times X$ are determined

by $\|(x, y)\|_{X \times X} = \|x\| + \|y\|$ and $\|(\lambda, x)\|_{\mathbf{F} \times X} = |\lambda| + \|x\|$. Therefore the desired continuities follow from

$$\begin{cases} \|(x+y) - (x_0+y_0)\| \leq \|x-x_0\| + \|y-y_0\| = \|(x, y) - (x_0, y_0)\|_{X \times X}; \\ \|\lambda x - \lambda_0 x_0\| \leq |\lambda - \lambda_0| \|x\| + |\lambda_0| \|x - x_0\| \leq \|(\lambda, x) - (\lambda_0, x_0)\|_{\mathbf{F} \times X} (\|x\| + |\lambda_0|). \end{cases}$$

8.5 It is clear that ℓ_0 is a linear subspace of ℓ_∞ . Note that $a = (1, 2^{-1}, 3^{-1}, \dots) \in \ell_\infty$. For every $n \in \mathbf{N}$, let $x_n = (1, 2^{-1}, 3^{-1}, \dots, n^{-1}, 0, 0, \dots) \in \ell_0$. Then

$$\|x_n - a\|_\infty = \|(0, \dots, 0, (n+1)^{-1}, (n+2)^{-1}, \dots)\|_\infty = (n+1)^{-1} \rightarrow 0.$$

It follows that $\{x_n\}_{n=1}^\infty$ converges in ℓ_∞ , but the limit a does not belong to ℓ_0 . Hence ℓ_0 is not closed in ℓ_∞ .

8.6 (i) By Hölder's inequality it follows that for $1 < p < q < \infty$,

$$\int_{[0,1]} |f|^p dm_{id} \leq \left(\int_{[0,1]} |f|^q dm_{id} \right)^{\frac{p}{q}},$$

and so that $\|f\|_{p, m_{id}, [0,1]}$ increases with p . Since $\|f\|_{\infty, m_{id}, [0,1]} < \infty$, the limit $a = \lim_{p \rightarrow \infty} \|f\|_{p, m_{id}, [0,1]}$ is finite. It is clear that

$$t(m_{id}(\{x \in [0, 1] : |f(x)| > t\}))^{\frac{1}{p}} \leq \|f\|_{p, m_{id}, [0,1]} \leq a$$

for all $t > 0$. Accordingly, if $m_{id}(\{x \in [0, 1] : |f(x)| > t\}) > 0$, then

$$\begin{cases} \lim_{p \rightarrow \infty} (m_{id}(\{x \in [0, 1] : |f(x)| > t\}))^{\frac{1}{p}} = 1; \\ m_{id}(\{x \in [0, 1] : |f(x)| > t\}) = 0 \text{ for } t > a, \end{cases}$$

and hence $a = \|f\|_{\infty, m_{id}, [0,1]}$.

(ii) It suffices to prove the result under $0 < \|f_1\|_{1, m_{id}, [0,1]} < \infty$ and $0 < \|f_2\|_{\infty, m_{id}, [0,1]} < \infty$. For the inequality, one has

$$\begin{aligned} \int_{[0,1]} |f_1 f_2| dm_{id} &= \int_{\{x \in [0,1] : |f_2(x)| \leq \|f_2\|_{\infty, m_{id}, [0,1]}\}} |f_1 f_2| dm_{id} \\ &\leq \|f_1\|_{1, m_{id}, [0,1]} \|f_2\|_{\infty, m_{id}, [0,1]}. \end{aligned}$$

As for the equality, note that $|f_2| \leq \|f_2\|_{\infty, m_{id}, [0,1]}$ holds m_{id} -a.e. on $[0, 1]$. So, if the equality is valid then

$$\int_{[0,1]} |f_1| (1 - |f_2| \|f_2\|_{\infty, m_{id}, [0,1]}^{-1}) dm_{id} = 0$$

and hence $|f_2(x)| = \|f_2\|_{\infty, m_{id}, [0,1]}$ for m_{id} -a.e. $x \in [0, 1]$. The reversed assertion is trivial.

8.7 It is clear that T is linear. Furthermore, if $f \in C[0, 1]$ then $|T(f)| = |f(0)| \leq \|f\|_{\infty}$, giving the boundedness of T with $\|T\| \leq 1$. If $f(x) = 1$ then $\|f\|_{\infty} = 1$ and $T(f) = 1$. This yields $\|T\| = 1$.

8.8 (i) Clearly,

$$\begin{cases} \|T(x)\|_{\infty} \leq \sup_{1 \leq k \leq n} \sum_{j=1}^n |a_{kj}| |x_j| \leq \|x\|_{\infty} \sup_{1 \leq k \leq n} \sum_{j=1}^{\infty} |a_{kj}|; \\ \|T\| \leq \sup_{1 \leq k \leq n} \sum_{j=1}^n |a_{kj}|. \end{cases}$$

To get the equality, just take $x_j = \text{sgn} a_{k_0j}$ where $\sum_{j=1}^n |a_{k_0j}| = \sup_{1 \leq k \leq n} \sum_{j=1}^n |a_{kj}|$. Then $\|x\|_{\infty} = 1$ and $y_k = \sum_{j=1}^n |a_{k_0j}|$. So

$$\|T\| \geq \|\{y_k\}_{k=1}^n\|_{\infty} = \sup_{1 \leq k \leq n} \left| \sum_{j=1}^n a_{kj} \text{sgn} a_{k_0j} \right| \geq \left| \sum_{j=1}^n a_{k_0j} \text{sgn} a_{k_0j} \right| = \sup_{1 \leq k \leq n} \sum_{j=1}^n |a_{kj}|,$$

giving the desired equality.

(ii) It follows from the definition of $\|\cdot\|_2$.

(iii) It is clear that for $f \in R^1[a, b]$,

$$\|T(f)\|_1 \leq \int_a^b \left(\int_a^x |f(t)| dt \right) dx \leq \int_a^b \left(\int_a^b |f(t)| dt \right) dx = (b-a) \|f\|_1,$$

and so that T is bounded with $\|T\| \leq b-a$. To see the equality, for any $n \in \mathbf{N}$ with $a + n^{-1} < b$ let

$$f_n(t) = \begin{cases} n, & t \in [a, a+n^{-1}] \\ 0, & t \in (a+n^{-1}, b] \end{cases}.$$

It is easy to check $\|f_n\|_1 = 1$ and

$$\|T(f_n)\|_1 = \int_a^{a+n^{-1}} n(x-a) dx + \int_{a+n^{-1}}^b dx = (b-a) - (2n)^{-1}.$$

So $\|T\| \geq \sup_{n \in \mathbf{N}} \|T(f_n)\|_1 = b-a$. Therefore $\|T\| = b-a$.

8.9 The boundedness is obvious. Since $T(f_1) = T(f_2)$ implies $f_1 = f_2$, we conclude that T is 1-1. But, T is not onto. In fact, if $T(f) = 1$ on $(0, 1)$ then the only possible candidate for f is $f(t) = t^{-1}$. It is clear that $f \in C(0, 1)$ with $\|f\|_2 = \infty$. From this it turns out that T is not invertible.

8.10 Here, suppose that the norm on \mathbf{R}^2 is $(2, 2)$ -norm.

Linear: $T(\alpha(x_1, y_1) + \beta(x_2, y_2)) = (\alpha x_1 + \beta x_2, 0) = \alpha T((x_1, y_1)) + \beta T((x_2, y_2))$.

Bounded: $\|T((x_1, x_2))\|_{2,2} = |x_1| \leq \|(x_1, x_2)\|_{2,2}$.

Not onto: $(x_1, 1)$ has no preimage under T .

An example: $S = (0, 1) \times (0, 1)$ is an open set of \mathbf{R}^2 , but $T(S) = \{(x_1, 0) : x_1 \in (0, 1)\}$ is not open set of \mathbf{R}^2 .

8.11 These results just follow from the definition.

8.12 (i) If $x \in X$ and $0 < \varepsilon < \|x\|_X$, then $x \in B_{\|x\|_X + \varepsilon}(0)$ and $x \notin B_{\|x\|_X - \varepsilon}(0)$. This readily implies the desired result.

(ii) Clearly, we have

$$p(\kappa x) = \inf\{\lambda > 0 : \kappa x \in \lambda K\} = \kappa \inf\{\kappa^{-1} \lambda > 0 : x \in \kappa^{-1} \lambda K\} = \kappa p(x).$$

Moreover, if $\varepsilon, \lambda, \mu$ are positive and satisfy $x = \lambda x_1$, $y = \mu y_1$, $\lambda < p(x) + 2^{-1}\varepsilon$ and $\mu < p(y) + 2^{-1}\varepsilon$ for some $x_1, y_1 \in K$, then

$$x + y = (\lambda + \mu) \left(\frac{\lambda}{\lambda + \mu} x_1 + \frac{\mu}{\lambda + \mu} y_1 \right) \in (\lambda + \mu) K.$$

So, $p(x + y) \leq \lambda + \mu < p(x) + p(y) + \varepsilon$ and hence $p(x + y) \leq p(x) + p(y)$.

(iii) Consider $L_0(\lambda x_0) = \lambda$ for $\lambda \in \mathbf{R}$ which is a linear functional on the linear subspace $\langle x_0 \rangle = \{\lambda x_0 : \lambda \in \mathbf{R}\}$. Since p is the Minkowski functional of K , we use the fact that $x_0 \notin K$ to obtain $\lambda x_0 \notin \lambda K$ for $\lambda \neq 0$ and thus $p(\lambda x_0) \geq \lambda = L_0(\lambda x_0)$. Consequently, there exists a linear extension L of L_0 to X obeying $L(x) \leq p(x)$ for any $x \in X$. When $x \in K$, we have $L(x) \leq p(x) \leq 1 = L_0(x_0)$. Note that $0 \in K^\circ$. So $\pm \varepsilon^{-1} y \in K$ for a given $\varepsilon > 0$ and any y with $\|y\|_X$ being small enough. Accordingly, $|L(y)| < \varepsilon$ for $\|y\|_X$ small enough, say $\|y\|_X < \delta$. Finally, the linearity of L gives that $|L(x_1) - L(x_2)| = |L(x_1 - x_2)| < \varepsilon$ whenever $\|x_1 - x_2\|_X < \delta$, namely, L is continuous.

8.13 Note that $E_1 - E_2 = \{a - b : a \in E_1, b \in E_2\}$ is convex and $0 \notin E_1 - E_2$. So, by hypothesis, there is an $x_0 \in (E_1 - E_2)^\circ$. Consequently, $-x_0 + E_1 - E_2$ is convex and contains 0 in its interior. Of course, $-x_0 \notin -x_0 + E_1 - E_2$. Thus, an application of Problem 8.12 (iii) yields a continuous linear functional L such that $L(-x_0 + a - b) \leq L(-x_0)$ for any $a \in E_1$ and $b \in E_2$. Consequently, $\sup_{a \in E_1} L(a) \leq \inf_{b \in E_2} L(b)$. Now, any $\alpha \in [\sup_{a \in E_1} L(a), \inf_{b \in E_2} L(b)]$ gives the two inclusions.

8.14 Let $f \in \mathcal{B}(X, \mathbf{F})$. If $\{x_n\}_{n=1}^\infty$ converges to $x \in X$ then $|f(x_n) - f(x)| \leq \|f\| \|x_n - x\|_X \rightarrow 0$ and hence $f(x) = 0$, i.e., $x \in f^{-1}(0)$. Thus $f^{-1}(0)$ is closed. Conversely, assume that $f^{-1}(0)$ is closed. If f is not in $\mathcal{B}(X, \mathbf{F})$, then there is a sequence of points $\{x_n\}_{n=1}^\infty$ in X such that $\|x_n\|_X = 1$ but $c_n = f(x_n) \rightarrow \infty$. Since f is linear, $f(x_n/c_n) = 1$ and consequently $f(x_n/c_n - x_1/c_1) = 0$; that is, $x_n/c_n - x_1/c_1 \in f^{-1}(0)$. Note that $x_n/c_n \rightarrow 0$ in X . So $x_n/c_n - x_1/c_1 \rightarrow -x_1/c_1$. Now, the assumption gives $x_1/c_1 \in f^{-1}(0)$, contradicting $f(x_1/c_1) = 1$. Therefore, $f \in \mathcal{B}(X, \mathbf{F})$.

8.15 Suppose $x_k = m_k + \lambda_k x \in M + \mathbf{F}x$ is convergent to $y \in X$. Then $\{\lambda_k\}_{k=1}^\infty$ is bounded – otherwise, there exists a subsequence λ_{k_j} such that $0 < |\lambda_{k_j}| \rightarrow \infty$ and hence $x + m_{k_j}/\lambda_{k_j} = x_{k_j}/\lambda_{k_j} \rightarrow 0$ in X ; but, M is closed, so we see that $x \in M$, contradicting the assumption $x \in X \setminus M$. Now the boundedness of $\{\lambda_k\}_{k=1}^\infty$ in \mathbf{F} yields a subsequence, say $\{\lambda_{k_j}\}_{j=1}^\infty$, which is convergent to a point $\lambda \in \mathbf{F}$. Consequently, $m_{k_j} = x_{k_j} - \lambda_{k_j} x \rightarrow y - \lambda x \in M$; this implies, $y \in M + \mathbf{F}x$. Therefore, $M + \mathbf{F}x$ is closed.

8.16 If $M = X$ then N is taken to be $\{0\}$. Now, suppose M is a closed proper subspace of X . For any $x \in X \setminus M$, let $L(y + \lambda x) = \lambda d(x, M)$ for all $y \in M$ and $\lambda \in \mathbf{F}$, then L is a linear functional defined on $M + \mathbf{F}x$ over \mathbf{F} . Clearly, $L(x) = d(x, M)$, $L(y) = 0$ for all $y \in M$, and $\|L\| = 1$. If L_x is the Hahn-Banach extension of L to X associated with x , then $L_x \in \mathcal{B}(X, \mathbf{F})$ with $\|L_x\| = 1$, $L_x(x) = d(x, M)$ and $M \subseteq N_{L_x} = \{z \in X : L_x(z) = 0\}$. Next, if M is finite dimensional, then it is closed. Moreover, by Corollary 8.4.4 (vii) for every $x \in X \setminus M$ there is a one-dimensional subspace Y_x of X such that $X = N_{L_x} + Y_x$ and $N_{L_x} \cap Y_x = \{0\}$. Accordingly, $M = \bigcap_{x \in X \setminus M} N_{L_x}$. To see this equation, assume that it is not true. By the first part of the argument, there exists an $x_0 \in \bigcap_{x \in X \setminus M} N_{L_x} \setminus M$ and hence an $L_{x_0} \in \mathcal{B}(X, \mathbf{F})$ such that $L_{x_0}(x_0) = d(x_0, M)$ and $M \subseteq N_{L_{x_0}}$. Note that $L_{x_0}(x) = 0$ for all $x \in X \setminus M$. So $L_{x_0}(x_0) = 0$, i.e., $x_0 \in M$ thanks to the closedness of M . This is a contradiction. Let $N = \bigcap_{x \in X \setminus M} Y_x$ which is closed. Since

$$X = \bigcap_{x \in X \setminus M} (N_{L_x} + Y_x) = \bigcap_{x \in X \setminus M} N_{L_x} + \bigcap_{x \in X \setminus M} Y_x,$$

we conclude $X = M + N$ and $M \cap N = \{0\}$.

8.17 The result follows from considering $p(x) = \inf\{f(y) : y \in M \text{ and } x \preceq y\}$ and adapting the argument for Theorem 8.4.3.

9 Banach Spaces via Operators and Functionals

9.1 If $\{f_j\}_{j=1}^\infty$ is a Cauchy sequence in $C^n[0, 1]$, then for any $\varepsilon > 0$ there exists $n_0 \in \mathbf{N}$ such that for all $x \in [0, 1]$,

$$m, n \geq n_0 \Rightarrow |f_m^{(k)}(x) - f_n^{(k)}(x)| \leq \|f_m - f_n\|_\infty < \varepsilon.$$

So for each integer $k \in [0, n]$, one has that $\{f_j^{(k)}(x)\}_{j=1}^\infty$ is a Cauchy sequence in \mathbf{R} . Therefore it converges to some real number $g_k(x)$ for every $x \in [0, 1]$; this defines a new function g_k such that $\{f_j^{(k)}\}_{j=1}^\infty$ converges to g_k pointwise on $[0, 1]$. Moreover, $\{f_j^{(k)}\}_{j=1}^\infty$ converges to g_k uniformly on $[0, 1]$, and that $g_k \in C[0, 1]$ – this follows readily from the last estimate when letting $m \rightarrow \infty$. In particular, $k = 0$ implies that $\{f_j\}_{j=1}^\infty$ converges uniformly on $[0, 1]$ to $g_0 = f$, and f is differentiable with $f' = g_1$. Furthermore, the same reasoning yields $f^{(k)} = g_k$ on $[0, 1]$. So, $f \in C^n[0, 1]$.

9.2 The result follows from the estimate for $m, n \in \mathbf{N}$,

$$\|x_m - x_n\|_X, \|y_m - y_n\|_Y \leq \|(x_m, y_m) - (x_n, y_n)\| \leq \|x_m - x_n\|_X + \|y_m - y_n\|_Y.$$

9.3 Consider $f_n(x) = x^n$ for each $n \in \mathbf{N}$. It is clear that $f_n \in C[0, 1]$ and the function

$$f(x) = \begin{cases} 1, & x = 1 \\ 0, & x \in [0, 1) \end{cases}$$

is not in $C[0, 1]$. But

$$\|f_n - f\|_p^p = \int_0^1 |x^n - f(x)|^p dx = \int_0^1 x^{pn} dx = (1 + pn)^{-1} \rightarrow 0.$$

9.4 It suffices to check the inequality $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$ for all $f, g \in L^1(\mathbf{Z}, \mathbf{F})$. Since $f, g \in L^1(\mathbf{Z}, \mathbf{F})$, it follows that

$$\sum_{m=-\infty}^{\infty} |f(n-m)| |g(m)| \leq \|f\|_1 \|g\|_1 < \infty$$

and so that for any $k, l \in \mathbf{Z}$,

$$\begin{aligned}
 \|f * g\|_1 &= \lim_{k \rightarrow \infty, l \rightarrow -\infty} \sum_{n=l}^k |f * g(n)| \\
 &\leq \lim_{k \rightarrow \infty, l \rightarrow -\infty} \sum_{n=l}^k \left(\lim_{j \rightarrow \infty} \sum_{m=-j}^j |f(n-m)| |g(m)| \right) \\
 &= \lim_{k \rightarrow \infty, l \rightarrow -\infty} \lim_{j \rightarrow \infty} \sum_{m=-j}^j \sum_{n=l}^k |f(n-m)| |g(m)| \\
 &\leq \|f\|_1 \lim_{j \rightarrow \infty} \sum_{m=-j}^j |g(m)| = \|f\|_1 \|g\|_1.
 \end{aligned}$$

9.5 Take $X = Y = \ell_0$ and equip it with the 2-norm: $\|x\|_2 = (\sum_{j=1}^{\infty} |x_j|^2)^{\frac{1}{2}}$. Here $x \in \ell_0$ if and only if $x \in \ell_{\infty}$ and it has only finitely many nonzero entries. Then X and Y are not complete. Now define $T_n(x) = (0, \dots, 0, nx_n, 0, \dots)$ for $x = (x_1, x_2, \dots)$. Then T_n is bounded with $\|T_n\| = n \rightarrow \infty$. On the other hand, if $x \in \ell_0$ then there is an $n_0 \in \mathbf{N}$ such that $n > n_0$, $x_n = 0$ and hence $\|T_n(x)\|_2 = n|x_n| = 0$ and if $n \leq n_0$ then $\|T_n(x)\|_2 = n|x_n| \leq n\|x\|_2 \leq n_0\|x\|_2$. Hence for each $x \in X$ we have $\sup_{n \in \mathbf{N}} \|T_n(x)\|_2 < \infty$. Clearly, the uniform bounded principle fails.

9.6 If $f(x) = (1+x^2)^{-1}$, then $f((-1, 1)) = (2^{-1}, 1]$ which is not open though $(-1, 1)$ is open.

9.7 (i) Since $|\sin x| \leq |x|$, we conclude that the integral is not less than

$$\int_0^{2\pi} 2x^{-1} |\sin(n+2^{-1})x| dx.$$

Note that

$$k\pi + 6^{-1}\pi \leq (n+2^{-1})x \leq k\pi + 3^{-1}\pi \Rightarrow |\sin(n+2^{-1})x| \geq 2^{-1}, \quad k \in \mathbf{N}.$$

So

$$\int_0^{2\pi} 2x^{-1} |\sin(n+2^{-1})x| dx \geq \sum_{k=0}^{2n} \left(\frac{\pi(k+3^{-1})}{n+2^{-1}} \right)^{-1} \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

(ii) It follows from a change of variable that

$$s_n(x) = (2\pi)^{-1} \int_{-x}^{2\pi-x} f(x+z) \left(\sum_{m=-n}^n e^{-imz} \right) dz.$$

This yields that $\overline{s_n(x)} = s_n(x)$ and hence $s_n(x)$ is real-valued. Note that $e^{ix} = \cos x + i \sin x$ and f is 2π -periodic. So

$$\begin{aligned} s_n(x) &= (2\pi)^{-1} \int_{-x}^{2\pi-x} f(x+z) \left(\sum_{m=-n}^n \cos mz \right) dz \\ &= (2\pi)^{-1} \int_{-x}^{2\pi-x} f(x+z) \left(1 + 2 \sum_{m=1}^n \cos mz \right) dz \\ &= (2\pi)^{-1} \int_{-x}^{2\pi-x} f(x+z) \left(\frac{\sin(n+2^{-1})z}{\sin 2^{-1}z} \right) dz \\ &= (2\pi)^{-1} \int_0^{2\pi} f(x+y) \left(\frac{\sin(n+2^{-1})y}{\sin 2^{-1}y} \right) dy. \end{aligned}$$

(iii) For any $f \in X$, we have

$$|T_n(f)| \leq (2\pi)^{-1} \|f\|_{\infty} \int_0^{2\pi} \left| \frac{\sin(n+2^{-1})x}{\sin 2^{-1}x} \right| dx,$$

and then

$$\|T_n\| \leq (2\pi)^{-1} \int_0^{2\pi} \left| \frac{\sin(n+2^{-1})x}{\sin 2^{-1}x} \right| dx.$$

To prove the equality actually holds, we may assume

$$q_n(x) = \frac{\sin(n+2^{-1})x}{\sin 2^{-1}x}$$

and $g_n(x) = \operatorname{sgn} q_n(x)$, that is,

$$g_n(x) = \begin{cases} 1 & , g_n(x) > 0 \\ 0 & , g_n(x) = 0 \\ -1 & , g_n(x) < 0 \end{cases}.$$

Then $|g_n(x)| = g_n(x)q_n(x)$. Though g_n is not continuous, for any $\varepsilon > 0$ there is a continuous function f_n such that

$$(2\pi)^{-1} \left| \int_0^{2\pi} (f_n(x) - g_n(x)) q_n(x) dx \right| < \varepsilon.$$

This can be easily realized since q_n is continuous on $[0, 2\pi]$. In fact, it is enough to use piecewise-defined segments to connect the discontinuous points of g_n so

that f_n is sufficiently close to g_n . Then $\|f_n\|_\infty = \max_{x \in [0, 2\pi]} |f_n(x)| = 1$, but

$$\begin{aligned} |\mathbb{T}_n(f_n)| &= (2\pi)^{-1} \int_0^{2\pi} f_n(x) q_n(x) dx \\ &= (2\pi)^{-1} \int_0^{2\pi} (f_n(x) - g_n(x)) q_n(x) dx + (2\pi)^{-1} \int_0^{2\pi} g_n(x) q_n(x) dx \\ &\geq (2\pi)^{-1} \left| \int_0^{2\pi} g_n(x) q_n(x) dx \right| - (2\pi)^{-1} \left| \int_0^{2\pi} (f_n(x) - g_n(x)) q_n(x) dx \right| \\ &\geq \int_0^{2\pi} |q_n(x)| dx - \varepsilon. \end{aligned}$$

Obviously, this implies

$$\|\mathbb{T}_n\| \geq (2\pi)^{-1} \int_0^{2\pi} \left| \frac{\sin(n+2^{-1})x}{\sin 2^{-1}x} \right| dx,$$

and so

$$\|\mathbb{T}_n\| = (2\pi)^{-1} \int_0^{2\pi} \left| \frac{\sin(n+2^{-1})x}{\sin 2^{-1}x} \right| dx.$$

(iv) It is clear that $\mathbb{T}_n(f) = s_n(0)$ for all $f \in X$. Moreover, if the Fourier series of a function $f \in X$ converges at 0, then $\{\mathbb{T}_n(f)\}_{n=1}^\infty$ is bounded since each element is just a partial sum of a convergent series. Thus if the Fourier series of f converges at 0 for all $f \in X$, then for each $f \in X$ the sequence $\{\mathbb{T}_n(f)\}_{n=1}^\infty$ is bounded. By the uniform bounded principle, this implies that $\{\|\mathbb{T}_n\|\}_{n=1}^\infty$ is bounded, contradicting (i) and (iii).

9.8 (i) Consider $f_n(x) = \sqrt{(x-2^{-1})^2 + n^{-2}}$. Then $\{f_n\}_{n=1}^\infty$ is convergent to the function $f(x) = |x-2^{-1}|$ uniformly on $[0, 1]$ since

$$|f_n(x) - |x-2^{-1}|| = \frac{n^{-2}}{\sqrt{(x-2^{-1})^2 + n^{-2}} + \sqrt{(x-2^{-1})^2}} \leq n^{-1}.$$

So, $\lim_{n \rightarrow \infty} \|f_n - f\|_\infty = 0$. However, f is not in $C^1[0, 1]$. Namely, X is not complete under the sup-norm.

(ii) To prove that $\mathbb{T} = d/dx$ is closed, let $\{x_n\}_{n=1}^\infty$ be a sequence in $C^1[0, 1]$ such that $x_n \rightarrow x$, $\mathbb{T}(x_n) \rightarrow y$. Since the convergence in $C[0, 1]$ means the uniform convergence, we conclude that $\mathbb{T}(x_n)$ converges to y uniformly on $[0, 1]$ and $y \in C[0, 1]$. Of course, $x \in C[0, 1]$ and $y(t) = x'(t)$; that is, \mathbb{T} is closed. But, it is not bounded since if $x_n(t) = \sin n\pi t$ then

$$\|\mathbb{T}(x_n)\|_\infty = \max_{t \in [0, 1]} |\cos n\pi t| n\pi = n\pi \rightarrow \infty.$$

9.9 The result follows from the related definitions.

9.10 (i) Assume the statement after the if and only if. If $L \in \mathcal{B}(C[0, 1], \mathbf{R})$ then by Theorem 9.3.4 there is a function g of bounded variation on $[0, 1]$ such that $L(f) = \int_0^1 f(x) dg(x)$ for all $f \in C[0, 1]$. Note that $\sup_{j \in \mathbf{N}} \|f_j\|_\infty < \infty$. So the dominated convergence theorem yields

$$\lim_{j \rightarrow \infty} L(f_j) = \lim_{j \rightarrow \infty} \int_0^1 f_j(x) dg(x) = \int_0^1 f(x) dg(x) = L(f),$$

namely, $\{f_j\}_{j=1}^\infty$ converges weakly to f .

Conversely, if $\{f_j\}_{j=1}^\infty$ converges weakly to f , then $\sup_{j \in \mathbf{N}} \|f_j\|_\infty < \infty$. For any $t \in [0, 1]$ define $L_t(f) = f(t)$ for $f \in C[0, 1]$. Clearly, $L_t \in \mathcal{B}(C[0, 1], \mathbf{R})$ and so

$$f_j(t) = L_t(f_j) \rightarrow L_t(f) = f(t) \text{ as } j \rightarrow \infty.$$

(ii) Assume the statement after the if and only if. Let $L \in \mathcal{B}(\ell_p, \mathbf{R})$. Then there exists a $y = (y_j)_{j=1}^\infty \in \ell_{p(p-1)^{-1}}$ such that

$$L(x) = \sum_{k=1}^\infty x_k y_k \text{ for all } x = \{x_k\}_{k=1}^\infty \in \ell_p.$$

Moreover, for any $\varepsilon > 0$ there exists an $n_0 \in \mathbf{N}$ such that $n > n_0$ implies

$$\sum_{k=1}^{n_0} |s_k^{(n)} - s_k| |y_k| < 2^{-1} \varepsilon$$

and

$$\sum_{k=n_0+1}^\infty |y_k|^{p(p-1)^{-1}} < \left(\frac{2^{-1} \varepsilon}{\|x\|_p + \sup_{j \in \mathbf{N}} \|x_j\|_p} \right)^{p(p-1)^{-1}}.$$

By Hölder's inequality,

$$\begin{aligned} & |L(x_j) - L(x)| \\ &= \left| \sum_{k=1}^\infty (s_k^{(j)} - s_k) y_k \right| \\ &\leq \sum_{k=1}^{n_0} |s_k^{(j)} - s_k| |y_k| + \sum_{k=n_0+1}^\infty |s_k^{(j)} y_k| + \sum_{k=n_0+1}^\infty |s_k y_k| \\ &\leq 2^{-1} \varepsilon + \left(\left(\sum_{k=n_0+1}^\infty |s_k^{(j)}|^p \right)^{\frac{1}{p}} + \left(\sum_{k=n_0+1}^\infty |s_k|^p \right)^{\frac{1}{p}} \right) \left(\sum_{k=n_0+1}^\infty |y_k|^{p(p-1)^{-1}} \right)^{(p-1)p^{-1}} \\ &\leq 2^{-1} \varepsilon + (\|x\|_p + \sup_{j \in \mathbf{N}} \|x_j\|_p) \left(\sum_{k=n_0+1}^\infty |y_k|^{p(p-1)^{-1}} \right)^{(p-1)p^{-1}} < \varepsilon. \end{aligned}$$

That is to say, $x_j \rightarrow x$ weakly.

Conversely, if $\{x_j\}_{j=1}^\infty$ converges weakly to x then the uniformly bounded principle gives $\sup_{j \in \mathbf{N}} \|x_j\|_p < \infty$. Let e_k be vector in ℓ_p for which the k th entry is 1 and others are 0. Then it produces an element $L_k \in \mathcal{B}(\ell_p, \mathbf{R})$ such that

$$x_j^{(k)} = L_k(x_j) \rightarrow L_k(x) = x_k \text{ as } j \rightarrow \infty.$$

This completes the argument.

9.11 Since X is separable, there is a countable dense subset $\{x_k\}_{k=1}^\infty$. Let $\{L_j\}_{j=1}^\infty$ be a sequence in M . Then $\sup_{j \in \mathbf{N}} \|L_j\| < \infty$ and so $\sup_{j \in \mathbf{N}} |L_j(x_k)| < \infty$ for each $k \in \mathbf{N}$. Using the Bolzano-Weierstrass property, we get a convergent subsequence $\{L_{j,k}(x_k)\}_{j=1}^\infty$ of $\{L_j(x_k)\}_{j=1}^\infty$. Now, $\{L_{n,n}\}_{n=1}^\infty$ is a subsequence of $\{L_j\}_{j=1}^\infty$ which is convergent on $\{x_k\}_{k=1}^\infty$. Thus the desired follows from Theorem 9.2.4 (ii).

9.12 Denote by J_1 the map $(\tilde{\cdot}) : X \rightarrow X^{**}$, and similarly, write J_2 for the map $(\tilde{\cdot}) : X^* \rightarrow X^{***}$. Suppose $J_1(X) = X^{**}$. For any $L \in X^{***}$ and $x \in X$ let $f(x) = L(J_1(x))$, then

$$(J_2(f))(J_1(x)) = J_1(x)(f) = f(x) = L(J_1(x)),$$

and hence $J_2(f) = L$, giving $J_2(X^*) = X^{***}$. Conversely, assume $J_2(X^*) = X^{***}$. If $J_1(X) \neq X^{**}$ then there is an $L \in X^{***}$ such that $L = 0$ on $J_1(X)$ but $\|L\| = 1$. The assumption then yields an $f \in X^*$ such that $J_2(f) = L$, and consequently

$$f(x) = J_1(x)(f) = (J_2(f))(J_1(x)) = L(J_1(x)) = 0 \text{ for } x \in X,$$

whence giving a contradiction $0 = \|J_2(f)\| = \|L\| = 1$. So $J_1(X) = X^{**}$.

9.13 (i) To prove that c_0 is a Banach space, it is enough to prove that c_0 is a closed subspace of ℓ_∞ . Suppose that $\{x^{(j)} = (x_k^{(j)})_{k=1}^\infty\}_{j=1}^\infty$ is a sequence in c_0 and convergent to $x = (x_k)_{k=1}^\infty$ in ℓ_∞ . Then for any $\varepsilon > 0$ there is a $j_0 \in \mathbf{N}$ such that $j > j_0$ implies $\|x^{(j)} - x\|_\infty < \varepsilon$. Note that $x^{(j_0+1)} \in c_0$. So there is a $k_0 \in \mathbf{N}$ such that $m, n > k_0$ implies $|x_m^{(j_0+1)} - x_n^{(j_0+1)}| < 3^{-1}\varepsilon$. Accordingly,

$$|x_m - x_n| \leq |x_m - x_m^{(j_0+1)}| + |x_m^{(j_0+1)} - x_n^{(j_0+1)}| + |x_n - x_n^{(j_0+1)}| < \varepsilon.$$

Namely, $\{x_k\}_{k=1}^\infty$ is a Cauchy sequence in \mathbf{R} and hence convergent to 0.

Next, we prove that c_0 is separable. For $n \in \mathbf{N}$ let

$$A_n = \{x \in c_0 : x = (r_1, r_2, \dots, r_n, 0, 0, \dots), r_j \in \mathbf{Q}\}.$$

Then the countable set $\cup_{n=1}^{\infty} A_n$ is a subset of c_0 . Now, given $x = (x_j)_{j=1}^{\infty} \in c_0$. Then for any $\varepsilon > 0$ there is an $n_0 \in \mathbf{N}$ such that $j > n_0$ implies $|x_j| < 2^{-1}\varepsilon$. Taking an $r_1, \dots, r_{n_0} \in \mathbf{Q}$ such that $\sup_{1 \leq j \leq n_0} |x_j - r_j| < 2^{-1}\varepsilon$, we obtain $y = (r_1, \dots, r_{n_0}, 0, 0, \dots) \in \cup_{n=1}^{\infty} A_n$ and

$$\|x - y\|_{\infty} \leq \sup_{1 \leq j \leq n_0} |x_j - r_j| + \sup_{j > n_0} |x_j| < \varepsilon.$$

Thus, $\cup_{n=1}^{\infty} A_n$ is a dense subset of c_0 , deriving the separability of c_0 .

Finally, if c_0 is reflexive then we would have $c_0 = \ell_0$ which is impossible.

(ii) It suffices to verify the necessity. For each $j \in \mathbf{N}$, let $e_j = (\underbrace{0, \dots, 0}_{j-1}, 1, 0, \dots)$.

Then $e_n \in c_0$. For any $L \in \mathcal{B}(c_0, \mathbf{R})$ set $y_j = L(e_j)$, we have

$$\begin{aligned} \sum_{j=1}^{\infty} |y_j| &= \sum_{j=1}^{\infty} L(e_j) \operatorname{sgn} L(e_j) \\ &= L\left(\sum_{j=1}^{\infty} e_j \operatorname{sgn} L(e_j)\right) \\ &\leq \|L\| \left\| \sum_{j=1}^{\infty} e_j \operatorname{sgn} L(e_j) \right\|_{\infty} \\ &\leq \|L\| \end{aligned}$$

and

$$L(x) = L\left(\sum_{j=1}^{\infty} x_j e_j\right) = \sum_{j=1}^{\infty} x_j y_j \quad \text{for } x = (x_k)_{k=1}^{\infty} \in c_0.$$

9.14 The weak closure clearly contains the norm closure. So it remains to check that if x is not in the norm closure of K then there exists a weak neighborhood of x which is disjoint from K . But, this follows right away from the convex separation theorem presented in Problem 8.13.

9.15 (i) It suffices to consider $A \neq X^*$. Note that for $f \in A$ and $g \in X^* \setminus A$ there is a point $x_0 \in X$ such that $f(x_0) \neq g(x_0)$ and hence for $0 < \varepsilon < |f(x_0) - g(x_0)|$ there are two disjoint neighborhoods

$$\begin{cases} O_f = \{F \in X^* : |F(x_0) - f(x_0)| < 2^{-1}\varepsilon\}; \\ O_g = \{F \in X^* : |F(x_0) - g(x_0)| < 2^{-1}\varepsilon\}. \end{cases}$$

So, a combination of the compactness of $A = \cup_{f \in A} O_f$ and the last part of the argument for Example 6.3.1 (iii) completes the proof.

(ii) Just follow the argument for Theorem 6.3.7 (i).

9.16 For any $x \in X$, suppose $T(x) \neq 0$ otherwise there is nothing to argue. Then

$$\|T(x)\|_Y = \sup_{\|f\|=1} |f \circ T(x)| \leq \sup_{\|f\|=1} \|f \circ T\| \|x\|_X,$$

thereby obtaining $T \in \mathcal{B}(X, Y)$.

9.17 Suppose X is compact. If $\{F_\alpha\}_{\alpha \in I}$ is a set of closed subsets of X and has the FIP, then any finite number of sets $\{F_j\}_{j=1}^n$ from $\{F_\alpha\}_{\alpha \in I}$ obeys $\cap_{j=1}^n F_j \neq \emptyset$, and hence $\cap_{\alpha \in I} F_\alpha \neq \emptyset$ – otherwise, $X = \cup_{\alpha \in I} (X \setminus F_\alpha)$. This yields a finite number of sets $\{X \setminus F_j\}_{j=1}^n$ such that $X = \cup_{j=1}^n X \setminus F_j$ thanks to X being compact. Consequently, $\cap_{j=1}^n F_j = \emptyset$ – a contradiction. Conversely, assume the statement after the if and only if. Suppose $X = \cup_{\alpha \in I} O_\alpha$ where $O_\alpha \subseteq X$ is open. Then $\cap_\alpha (X \setminus O_\alpha) = \emptyset$, and hence by the contrapositive statement of the assumption it follows that there is a finite number of sets $\{X \setminus O_j\}_{j=1}^n$ such that $\cap_{j=1}^n (X \setminus O_j) = \emptyset$ and consequently $X = \cup_{j=1}^n O_j$. Accordingly, X is compact.

9.18 Since $\mathcal{B}_c(X, Y)$ is a closed linear subspace of the Banach space $\mathcal{B}(X, Y)$, it is a Banach space, too.

9.19 The result follows from a modification of the argument for Example 9.5.7 – in fact, the proof does not employ the continuity of $k(\cdot, \cdot)$ on the region $\{(x, y) \in \mathbb{R}^2 : x < y, a \leq x, y \leq b\}$.

9.20 (i) Consider a bounded sequence $\{x_j\}_{j=1}^\infty$ in X . Since T is compact, the sequence $\{T(x_j)\}_{j=1}^\infty$ has a convergent subsequence, say, $\{T(x_{j_k})\}_{k=1}^\infty$ which is convergent to y in X . Then

$$\|ST(x_{j_k}) - S(y)\| \leq \|S\| \|T(x_{j_k}) - y\| \rightarrow 0 \text{ as } k \rightarrow \infty.$$

This shows that ST is compact. Note that $\{S(x_j)\}_{j=1}^\infty$ is bounded. So $\{TS(x_j)\}_{j=1}^\infty$ has a convergent subsequence, showing that TS is compact. Similarly, ST is compact, too.

(ii) For any $n \in \mathbb{N}$, let $B_n^X \subseteq X$ be the open ball centered at 0 with radius n . Then $T(X) = \cup_{n=1}^\infty T(B_n^X)$. Since $\overline{T(B_n^X)}$ is compact, $\overline{T(B_n^X)}$ is separable by Theorem 6.3.12 and hence $T(X)$ is separable.

9.21 (i) For any bounded set $B \subseteq X$, we have $\overline{T(B)} = \{0\}$ and so T is compact.

(ii) Assume that T is invertible. Then the identity operator $I = T^{-1}T$ is compact on X . But since X is infinite dimensional, the unit sphere S of X is not compact and consequently $\overline{I(S)} = S$ is not compact, i.e., I is not compact on X , a contradiction. Thus, T is not invertible.

(iii) Note that $T(X)$ is a closed linear subspace of the Banach space X . If T is compact then the closed unit ball of $T(X)$ is compact and hence $T(X)$ must be finite dimensional thanks to Theorem 8.2.5 (ii), contradicting the assumption that $T(X)$ is infinite dimensional.

10 Hilbert Spaces and Their Operators

10.1 It is easy to check that the inner product $\langle f, g \rangle = \int_{-1}^1 f(x) \overline{g(x)} dx$ satisfies (a), (b) and (c) of Definition 5.1.1, and so $C([-1, 1], \mathbf{C})$ is an inner product space. Since the norm equipped with $C([-1, 1], \mathbf{C})$ is the 2-norm: $\|f\|_2 = \sqrt{\int_{-1}^1 |f(x)|^2 dx}$, we can conclude that this space is not complete. In fact, for each $j \in \mathbf{N}$ let

$$f_j(x) = \begin{cases} -1, & x \in [-1, 0) \\ jx, & x \in [-j^{-1}, j^{-1}] \\ 1, & x \in [j^{-1}, 1] \end{cases},$$

then $f_j \in C([-1, 1], \mathbf{C})$. Observe that for each $m, n \in \mathbf{N}$ satisfying $m > n$, one has

$$\begin{aligned} \|f_m - f_n\|_2^2 &= 2 \left(\int_0^{m^{-1}} (mx - nx)^2 dx + \int_{m^{-1}}^{n^{-1}} (1 - nx)^2 dx \right) \\ &= 2 \left(\frac{(m-n)^2}{3m^3} + \frac{1}{3n} - \frac{1}{m} + \frac{n}{m^2} - \frac{n^2}{3m^3} \right) \\ &< \frac{6}{m} + \frac{1}{n} < \frac{7}{n} \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

This means that (f_j) is a Cauchy sequence in $C([-1, 1], \mathbf{C})$. On the other hand, (f_j) converges pointwise to

$$f(x) = \begin{cases} -1, & x \in [-1, 0) \\ 0, & x = 0 \\ 1, & x \in (0, 1] \end{cases},$$

which does not belong to $C([-1, 1], \mathbf{C})$. Therefore, $C([-1, 1], \mathbf{C})$ is not a Hilbert space.

10.2 It is clear that $\langle x, y \rangle = \sum_{j=1}^{\infty} x_j \overline{y_j}$ is an inner product defined on $\ell_2(\mathbf{N}, \mathbf{C})$. It is not hard to see that $\ell_2(\mathbf{N}, \mathbf{C})$ is complete under the norm $\|x\|_2 = \sqrt{\langle x, x \rangle}$. So, $\ell_2(\mathbf{N}, \mathbf{C})$ is a Hilbert space over \mathbf{C} . To obtain that ℓ_p is not a Hilbert space, one takes $x = (1, 1, 0, 0, \dots)$ and $y = (1, -1, 0, 0, \dots)$. Then $x, y \in \ell_p$ and $\|x\|_p = \|y\|_p = 2^{\frac{1}{p}}$ but $\|x + y\|_p = \|x - y\|_p = 2$. Hence the parallelogram law fails at $p \neq 2$; that is to say, if $p \neq 2$ then $\|\cdot\|_p$ does not induce an inner product, and hence ℓ_p is not a Hilbert space over \mathbf{C} .

10.3 (i) When X is a real inner product space, the equivalence follows from $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle$. But, if X is an inner product space over \mathbf{C} , then the equivalence fails since $\|x\| = 1$ and $y = ix$ give $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ and $\langle x, y \rangle = -i \neq 0$.

(ii) This follows from $\|x + \beta y\|^2 = \|x\|^2 + \|y\|^2 + \overline{\beta}\langle x, y \rangle + \beta\overline{\langle x, y \rangle} + |\beta|^2\|y\|^2$.

10.4 Obviously, $M \subseteq (M^\perp)^\perp$. To get the reverse inclusion, let $x \in (M^\perp)^\perp$. Then the projection theorem implies that there exist $y \in M \subseteq (M^\perp)^\perp$ and $z \in M^\perp$ such that $x = y + z$. Because $x \in (M^\perp)^\perp$ and $(M^\perp)^\perp$ is a linear space, one has $z = x - y \in (M^\perp)^\perp$ and so $z \in M^\perp \cap (M^\perp)^\perp = \{0\}$, i.e., $z = 0$ and $x = y$. This gives $(M^\perp)^\perp \subseteq M$.

10.5 (i) Given $g \in M^\perp$. Then for any $f \in M$ one has

$$0 = \int_{[-1, 1]} f \overline{g} dm_{id} = \int_{[0, 1]} f \overline{g} dm_{id}.$$

If f is taken to respectively be 0 a.e. on $[-1, 0]$ and g a.e. on $(0, 1]$, then $\int_0^1 |g(x)|^2 dx = 0$ and hence $g = 0$ a.e. on $(0, 1]$. It follows that $M^\perp = \{g \in L^2([-1, 1], \mathbf{C}) : g(x) = 0 \text{ m}_{id} \text{ a.e. on } (0, 1]\}$.

(ii) Any function $f \in L^2([-1, 1], \mathbf{C})$ can be written as $f_{\text{even}} + f_{\text{odd}}$ where

$$f_{\text{even}}(x) = \frac{f(x) + f(-x)}{2} \quad \text{and} \quad f_{\text{odd}}(x) = \frac{f(x) - f(-x)}{2}.$$

Clearly, $f_{\text{even}} \in M_{\text{even}}$ and $f_{\text{odd}} \in M_{\text{odd}}$. Moreover, if $f \in M_{\text{even}} \cap M_{\text{odd}}$, then $f(x) = -f(x)$ and hence $f(x) = 0$. In other words, $M_{\text{even}} \cap M_{\text{odd}} = \{0\}$. This implies the desired direct sum decomposition.

10.6 (i) This assertion follows from

$$(2\pi)^{-1} \int_{-\pi}^{\pi} e^{i(n-m)x} dx = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n \neq m. \end{cases}$$

(ii) First of all, a calculation gives

$$\begin{cases} \int_{-\pi}^{\pi} 1 dx = 2\pi; \\ \int_{-\pi}^{\pi} \cos^2 nx dx = \int_{-\pi}^{\pi} \sin^2 nx dx = \pi; \\ \int_{-\pi}^{\pi} \cos nx dx = \int_{-\pi}^{\pi} \sin nx dx = 0; \\ \int_{-\pi}^{\pi} \cos mx \cos nx dx = \int_{-\pi}^{\pi} \sin mx \sin nx dx = 0 \text{ for all } m \neq n, m, n \in \mathbf{N}; \\ \int_{-\pi}^{\pi} \sin mx \cos nx dx = 0 \text{ for all } m, n \in \mathbf{N}. \end{cases}$$

Now, let $\{e_j\}_{j=1}^{\infty}$ be this orthonormal set. According to the definition of orthonormal basis, we must prove that $\langle f, e_j \rangle = 0$ implies $f = 0$ a.e. on $[-\pi, \pi]$. To do so, let us first consider the case that f is continuous and real-valued. If $f \neq 0$ a.e. on $[-\pi, \pi]$, then there is an $x_0 \in [-\pi, \pi]$ at which $|f|$ achieves a maximum, and we may assume $f(x_0) > 0$. Thus, there is a $\delta > 0$ such that $f(x) > 2^{-1}f(x_0)$ for $x \in (x_0 - \delta, x_0 + \delta)$. If $g(x) = 1 + \cos(x_0 - x) - \cos \delta$, then $1 < g(x)$ for $x \in (x_0 - \delta, x_0 + \delta)$ and $|g(x)| \leq 1$ for $x \in [-\pi, \pi] \setminus (x_0 - \delta, x_0 + \delta)$. Note that $\langle f, e_j \rangle = 0$. So for any $n \in \mathbf{N}$,

$$\begin{aligned} 0 &= \langle f, g^n \rangle = \int_{-\pi}^{\pi} f(x) g^n(x) dx \\ &= \int_{-\pi}^{x_0 - \delta} f(x) g^n(x) dx + \int_{x_0 - \delta}^{x_0 + \delta} f(x) g^n(x) dx + \int_{x_0 + \delta}^{\pi} f(x) g^n(x) dx. \end{aligned}$$

Using the properties of g above, we obtain

$$\begin{cases} \left| \int_{-\pi}^{x_0 - \delta} f(x) g^n(x) dx \right| \leq 2\pi f(x_0); \\ \left| \int_{x_0 + \delta}^{\pi} f(x) g^n(x) dx \right| \leq 2\pi f(x_0); \\ \int_{x_0 - \delta}^{x_0 + \delta} f(x) g^n(x) dx \geq \int_a^b f(x) g^n(x) dx, \end{cases}$$

for all $[a, b] \subseteq (x_0 - \delta, x_0 + \delta)$. Since g is continuous on $[a, b]$, we can conclude that g achieves a minimum value, $\kappa > 1$, there. This implies

$$4\pi f(x_0) \geq \int_a^b f(x) g^n(x) dx \geq 2^{-1} f(x_0) \kappa^n (b - a) \rightarrow \infty \text{ as } n \rightarrow \infty.$$

This is a contradiction. Thus, $f = 0$ on $[-\pi, \pi]$.

If f is continuous but not real-valued, then our hypothesis gives that for all $k = 0, \pm 1, \pm 2, \dots$,

$$\int_{-\pi}^{\pi} f(x) e^{-ikx} dx = 0 \quad \text{and} \quad \int_{-\pi}^{\pi} \overline{f(x)} e^{-ikx} dx = 0.$$

Hence

$$\int_{-\pi}^{\pi} \Re f(x) e_j(x) dx = 0 \quad \text{and} \quad \int_{-\pi}^{\pi} \Im f(x) e_j(x) dx = 0.$$

By the first part, we get that $\Re f(x) = 0 = \Im f(x)$ and so $f = 0$ on $[-\pi, \pi]$.

Finally, we no longer assume that f is continuous. But, f generates a continuous function $F(x) = \int_{-\pi}^x f(t) dt$. An integration by parts yields $\int_{-\pi}^{\pi} F(x) \sin(kx) dx = 0$. Similarly, $\int_{-\pi}^{\pi} F(x) \cos(kx) dx = 0$. This infers that F and $F - C$ for every constant C , is orthogonal to each of the non-constant members of $\{e_j\}$. For $(2\pi)^{-2^{-1}}$, let $C_0 = (2\pi)^{-1} \int_{-\pi}^{\pi} F(x) dx$. Then $\langle F - C_0, e_j \rangle = 0$ for all $j \in \mathbb{N}$. Since F is continuous, we conclude from the first two parts that $F - C_0 = 0$. Of course, $f = F' = 0$ a.e. on $[-\pi, \pi]$.

10.7 (i) From $\langle T(f), g \rangle = \langle f, T^*(g) \rangle$ we have

$$\int_{[-\pi, \pi]} \phi f \bar{g} dm_{id} = \int_{[-\pi, \pi]} f \overline{T^*(g)} dm_{id}$$

and so $T^*(g) = \bar{\phi} g$ by the uniqueness of T^* .

(ii) It is clear that if ϕ is real-valued, then $\bar{\phi} = \phi$ and hence $T^* = T$.

(iii) If $|\phi| = 1$ a.e. on $[-\pi, \pi]$ then T is unitary; If $\phi \geq 0$ m_{id} -a.e. on $[-\pi, \pi]$ then $\langle T(f), f \rangle \geq 0$ and hence T is positive; If $\phi = 1$ or 0 m_{id} -a.e. on $[-\pi, \pi]$ then $T^2(f) = \phi^2 f = \phi f = T(f)$ and hence T is projection.

10.8 If $x \in N(T)$ then $T(x) = 0$ and hence $\langle x, T^*(y) \rangle = \langle T(x), y \rangle = 0$ for all $y \in X$, $T^*y \in R(T^*)$. This gives $x \in R(T^*)^\perp$. Conversely, if $x \in R(T^*)^\perp$, then $\langle x, T^*(y) \rangle = \langle T(x), y \rangle = 0$ for any $y \in X$, and hence $T(x) = 0$, i.e., $x \in N(T)$. Therefore $N(T) = R(T^*)^\perp$. Replacing T by T^* , we get $N(T^*) = R(T)^\perp$ and $N(T^*)^\perp = (R(T)^\perp)^\perp$.

Since $R(T) \subseteq \overline{R(T)}$, by the continuity of $\langle \cdot, \cdot \rangle$ and the closedness of $\overline{R(T)}$ we obtain $\overline{R(T)} = (\overline{R(T)}^\perp)^\perp = (R(T)^\perp)^\perp$. Thus $N(T^*)^\perp = \overline{R(T)}$. Replacing T^* by T , we also obtain $\overline{R(T^*)} = N(T)^\perp$.

10.9 (i) Consider $(T - \lambda I)(x) = y$. Thus $x(t) = y(t)/(t - \lambda)$ provided that $t \neq \lambda$ for $t \in [0, 2\pi]$. So there is a unique solution $x \in C[0, 2\pi]$ except when $t = \lambda$ for $t \in [0, 2\pi]$ and so $\sigma(T) = [0, 2\pi]$.

(ii) The argument is similar to that for (i).

10.10 (i) For $\lambda \in \mathbf{C}$ with $|\lambda| < 1$ let $x_\lambda = (1, \lambda, \dots, \lambda^n, \dots)$. Then $x_\lambda \in \ell_2(\mathbf{N}, \mathbf{C})$ and $T(x_\lambda) = (\lambda, \lambda^2, \dots, \lambda^n, \dots) = \lambda x_\lambda$. This means $\{\lambda \in \mathbf{C} : |\lambda| \leq 1\} \subseteq \sigma(T)$. On the other hand, $\|T(x)\|_2 \leq \|x\|_2$ for any $x \in \ell_2(\mathbf{N}, \mathbf{C})$, and thus $\|T\| \leq 1$, giving $\sigma(T) \subseteq \{\lambda \in \mathbf{C} : |\lambda| \leq \|T\|\} \subseteq \{\lambda \in \mathbf{C} : |\lambda| \leq 1\}$. Therefore, $\sigma(T) = \{\lambda \in \mathbf{C} : |\lambda| \leq 1\}$.

(ii) Refer to the argument for (i).

10.11 The formula follows from $(\lambda I - T)^* = \bar{\lambda} I - T^*$.

10.12 (i) This follows from that fact that $\langle T(x), x \rangle, \lambda \in \mathbf{R}$ ensures

$$\|T(x) - \lambda x\|^2 = (\lambda - \langle T(x), x \rangle)^2 + \|T(x)\|^2 - \langle T(x), x \rangle^2.$$

(ii) By Theorem 10.5.8 we can write $x = x_0 + \sum_{k=1}^{\infty} P_k(x)$ and $T(x) = \sum_{k=1}^{\infty} \lambda_k P_k(x)$ where $x_0 \in N(T)$ and $\{\lambda_k\}_{k=1}^{\infty}$ are those nonzero eigenvalues of T . Then

$$T(x) - \langle T(x), x \rangle x = \sum_{k=1}^{\infty} (\lambda_k - \langle T(x), x \rangle) P_k(x) - \langle T(x), x \rangle x_0.$$

Since $\sigma(T)$ is closed, we can choose a point $\lambda \in \sigma(T)$ such that it is nearest to $\langle T(x), x \rangle$, we use (i) and the Pythagorean property to get

$$\begin{aligned} \|T(x) - \langle T(x), x \rangle x\|^2 &= \sum_{k=1}^{\infty} (\lambda_k - \langle T(x), x \rangle)^2 \|P_k(x)\|^2 + \langle T(x), x \rangle^2 \|x_0\|^2 \\ &\geq |\lambda - \langle T(x), x \rangle|^2 \left(\sum_{k=1}^{\infty} \|P_k(x)\|^2 + \|x_0\|^2 \right) \\ &= |\lambda - \langle T(x), x \rangle|^2 \|x\|^2 \\ &= |\lambda - \langle T(x), x \rangle|^2. \end{aligned}$$

10.13 Let $c = \int_0^1 e^y f(y) dy$ and f obey $T(f) = \lambda f$. Then $ce^x = \lambda f(x)$. If $\lambda \neq 0$ then $f(x) = (c/\lambda)e^x$ and hence $c = \int_0^1 e^y (c/\lambda) e^y dy$, yielding $\lambda = (e^2 - 1)/2$ and $f(x) = c_0 e^x$ where c_0 is any nonzero constant. If $\lambda = 0$ then $\int_0^1 e^x f(x) dx = 0$ and hence eigenvectors are all the nonzero elements in $\{e^x\}^\perp$.

10.14 Since $T_1 = (T + T^*)/2$ and $T_2 = (T - T^*)/(2i)$, we use $TT^* = T^*T$ to obtain

$$T_1^2 + T_2^2 = 2^{-2}(T + T^*)(T + T^*) - 2^{-2}(T - T^*)(T - T^*) = TT^*$$

thereby achieving $\|T_1^2 + T_2^2\| = \|TT^*\| = \|T\|^2$ and $\|T^2\|^2 = \|(T^2)^*T^2\| = \|TT^*\|^2 = \|T\|^4$, as desired.

10.15 (i) This follows from $\|(T^* - \bar{\lambda}I)(x)\|^2 = \|(T - \lambda I)(x)\|^2$.

(ii) and (iii) These follow from a simple calculation via the definition.

10.16 Use $\|T(x_m) - T(x_n)\|^2 \leq \|T^*T(x_m) - T^*T(x_n)\| \|x_m - x_n\|$.

10.17 Given $j \in \mathbf{N}$, set $e_j = (\underbrace{0, \dots, 0}_{j-1}, 1, 0, \dots)$. For $x = \{x_j\}_{j=1}^\infty \in \ell_2(\mathbf{N}, \mathbf{C})$ and $n \in \mathbf{N}$ let $T_n(x) = \sum_{j=1}^n (\sum_{k=1}^\infty x_k a_{jk}) e_j$. Then T_n has finite rank and hence it is compact. Note that

$$\|(T - T_n)(x)\|_2^2 = \sum_{j=n+1}^\infty \left| \sum_{k=1}^\infty x_k a_{jk} \right|^2 \leq \left(\sum_{j=n+1}^\infty \sum_{k=1}^\infty |a_{jk}|^2 \right) \|x\|_2^2$$

thanks to the Cauchy-Schwarz inequality. So

$$\|T - T_n\| \leq \sqrt{\sum_{j=n+1}^\infty \sum_{k=1}^\infty |a_{jk}|^2} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

This implies the compactness of T .

10.18 (i) Note that

$$\begin{aligned} K(f)(x) &= \int_{[0,1]} \left(\sum_{k=1}^n g_k(x) f_k(y) \right) f(y) dm_{id}(y) \\ &= \sum_{k=1}^n \left(\int_{[0,1]} f_k f dm_{id} \right) g_k(x). \end{aligned}$$

So, the range of K is finite-dimensional, and consequently, K is compact.

(ii) Just use the formula in (i).

(iii) In this case, we have

$$h_{j,k} = \begin{cases} 0 & , j \neq k \\ \int_{[0,1]} g_j^2 dm_{id} & , j = k \end{cases}.$$

Moreover, $(h_{j,j} - \lambda)c_j$ for $j = 1, 2, \dots, n$. This means that $\{h_{j,j}\}_{j=1}^n$ is the set of all nonzero eigenvalues, and g_j is the eigenvector corresponding to $h_{j,j}$ for each j . It is clear that if $M = \text{span}\{g_j\}_{j=1}^n$ then any nonzero element in M^\perp is an eigenvector corresponding to 0.

(iv) Since $\cos \pi(y+x) = \cos \pi y \cos \pi x + (i \sin \pi y)(i \sin \pi x)$, we can take $g_1(x) = f_1(x) = \cos(\pi x)$ and $g_2(x) = f_2(x) = i \sin(\pi x)$, obtaining $\langle g_1, g_2 \rangle = 0$ and $h_{1,1} = 2^{-1}$; $h_{2,2} = -2^{-1}$. Accordingly, K has two nonzero eigenvalues $\lambda_1 = 2^{-1}$ and $\lambda_2 = -2^{-1}$ for which the corresponding eigenvectors are $c \cos \pi x$ and $c \sin \pi x$ respectively. Of course, the eigenvectors corresponding to 0 are the nonzero functions in $\{\cos \pi x, \sin \pi x\}^\perp$.

10.19 Assuming that $\lim_{j \rightarrow \infty} \|\mathsf{T}(e_j)\| \rightarrow 0$ is not true, we use the compactness of T to get a subsequence $\{e_{j_k}\}_{k=1}^\infty$ and a nonzero $x \in X$ such that $\|\mathsf{T}(e_{j_k}) - x\| < k^{-1}$ holds for any $k \in \mathbf{N}$. Letting $x_n = \sum_{k=1}^n k^{-1} e_{j_k}$, we find that $\{x_n\}_{n=1}^\infty$ is convergent in X and $\{\mathsf{T}(x_n)\}_{n=1}^\infty$ is bounded since $\sum_{k=1}^n k^{-2} < \infty$ and $\{k^{-1} e_{j_k}\}_{k=1}^\infty$ is orthonormal, but also (by the triangle inequality)

$$\begin{aligned} \|\mathsf{T}(x_n)\| &= \left\| \sum_{k=1}^n k^{-1} \mathsf{T}(e_{j_k}) \right\| \\ &= \left\| \sum_{k=1}^n k^{-1} (\mathsf{T}(e_{j_k}) - x) + \left(\sum_{k=1}^n k^{-1} \right) x \right\| \\ &\geq \|x\| \sum_{k=1}^n k^{-1} - \sum_{k=1}^n \|\mathsf{T}(e_{j_k}) - x\| \\ &\geq \|x\| \sum_{k=1}^n k^{-1} - \sum_{k=1}^n k^{-2} \rightarrow \infty, \end{aligned}$$

contradicting the boundedness of $\{\mathsf{T}(x_n)\}_{n=1}^\infty$.

10.20 (i) Suppose that $\{f_k\}_{k=1}^\infty$ is an orthonormal basis of Y . Then the result follows from

$$\begin{aligned} \sum_{j=1}^\infty \|\mathsf{T}(e_j)\|_Y^2 &= \sum_{j=1}^\infty \left(\sum_{k=1}^\infty |\langle \mathsf{T}(e_j), f_k \rangle_Y|^2 \right) \\ &= \sum_{k=1}^\infty \left(\sum_{j=1}^\infty |\langle e_j, \mathsf{T}^*(f_k) \rangle_X|^2 \right) \\ &= \sum_{k=1}^\infty \|\mathsf{T}^*(f_k)\|_X^2. \end{aligned}$$

(ii) For each $k \in \mathbf{N}$ and $x = \sum_{j=1}^{\infty} \langle x, e_j \rangle_X e_j \in X$ let

$$T_k(x) = T\left(\sum_{j=1}^k \langle x, e_j \rangle_X e_j\right) = \sum_{j=1}^k \langle x, e_j \rangle_X T(e_j).$$

Then T_k is a compact operator from X to Y . This fact, plus

$$\begin{aligned} \|(T_k - T)(x)\|_Y &\leq \sum_{j=k+1}^{\infty} |\langle x, e_j \rangle_X| \|T(e_j)\|_Y \\ &\leq \left(\sum_{j=k+1}^{\infty} |\langle x, e_j \rangle_X|^2 \right)^{\frac{1}{2}} \left(\sum_{j=k+1}^{\infty} \|T(e_j)\|_Y^2 \right)^{\frac{1}{2}} \\ &\leq \|x\|_X \left(\sum_{j=k+1}^{\infty} \|T(e_j)\|_Y^2 \right)^{\frac{1}{2}}, \end{aligned}$$

derives

$$\|T_k - T\| \leq \left(\sum_{j=k+1}^{\infty} \|T(e_j)\|_Y^2 \right)^{\frac{1}{2}} \rightarrow 0 \text{ as } k \rightarrow \infty,$$

whence the compactness of T .

(iii) Choosing any orthonormal basis $\{e_j\}_{j=1}^{\infty}$ in $LRS_{id}^2((c, d), \mathbf{C})$, we get

$$T(e_j)(x) = \int_{(c,d)} k(x, y) e_j(y) dm_{id}(y) = \langle k(x, \cdot), \overline{e_j} \rangle_{(c,d)}$$

where the right-hand inner product is defined on $LRS_{id}^2((c, d), \mathbf{C})$. Thus,

$$\begin{aligned} \sum_{j=1}^{\infty} \|T(e_j)\|^2 &= \lim_{k \rightarrow \infty} \sum_{j=1}^k \|T(e_j)\|^2 \\ &= \lim_{k \rightarrow \infty} \sum_{j=1}^k \int_{(a,b)} |\langle k(x, \cdot), \overline{e_j} \rangle_{(c,d)}|^2 dm_{id}(x) \\ &= \lim_{k \rightarrow \infty} \int_{(a,b)} \sum_{j=1}^k |\langle k(x, \cdot), \overline{e_j} \rangle_{(c,d)}|^2 dm_{id}(x) \\ &\leq \lim_{k \rightarrow \infty} \int_{(a,b)} \int_{(c,d)} |k(x, y)|^2 dm_{id}(y) dm_{id}(x) < \infty. \end{aligned}$$

10.21 (i) Let $M = \sup_{k \in \mathbf{N}} |c_k|$. If $M < \infty$ then $\|T(x)\|_Y^2 \leq M^2 \|x\|_X$ for any $x \in X$ and hence T is bounded. Conversely, if T is bounded but $M = \infty$, then for any

$n \in \mathbf{N}$ there is c_n such that $|c_n| \geq n$, and hence $\|\mathbf{T}(e_n)\|_Y = \|c_n f_n\|_Y = |c_n| \geq n$, contradicting the boundedness of \mathbf{T} .

(ii) Let $\lim_{k \rightarrow \infty} |c_k| = 0$. For each $n \in \mathbf{N}$ and $x \in X$ define $\mathbf{T}_n(x) = \sum_{j=1}^n c_k \langle x, e_k \rangle_X f_k$. Then this operator is compact. It is easy to see that $\lim_{n \rightarrow \infty} \|\mathbf{T}_n - \mathbf{T}\| = 0$. So \mathbf{T} is compact. Conversely, if \mathbf{T} is compact but $(|c_k|)_{k=1}^\infty$ does not approach 0, then there is an $\epsilon_0 > 0$ and a subsequence $\{c_{k_j}\}_{j=1}^\infty$ such that $|c_{k_j}| \geq \epsilon_0$ for all $j \in \mathbf{N}$. Consequently, $j, l \in \mathbf{N}$ with $j \neq l$ implies

$$\|\mathbf{T}(e_{k_j}) - \mathbf{T}(e_{k_l})\|_Y^2 = \|c_{j_k} f_{j_k} - c_{j_l} f_{j_l}\|_Y^2 = |c_{j_k}|^2 + |c_{j_l}|^2 \geq 2\epsilon_0^2$$

and then no subsequence of $\{\mathbf{T}(e_{k_j})\}_{j=1}^\infty$ is convergent, i.e., \mathbf{T} is not compact, a contradiction.

(iii) This follows from $\mathbf{T}(e_j) = c_j f_j$.

(iv) Since $f_k \in R(\mathbf{T})$ for any $k \in \mathbf{N}$ with $c_k \neq 0$, the desired equivalence follows from the linear independence of $\{f_k\}_{k=1}^\infty$.

References

- [AmT] Ambrosio, L., Tilli, P.: *Topics on Analysis in Metric Spaces*, Oxford University Press, 2004.
- [Apo] Apostol, T. M.: *Mathematical Analysis*, Addison-Wesley Publishing Company, Inc. 1964.
- [Bol] Bollobás, B.: *Linear Analysis*, Cambridge University Press, 1990.
- [Bur] Burk, F.: *Lebesgue Measure and Integration*, John Wiley & Sons, Inc., 1998.
- [Cain] Cain, G. L.: *Introduction to General Topology*, Addison-Wesley Publishing Company, 1993.
- [CarB] Carter, M., Brunt, B.V.: *The Lebesgue-Stieltjes Integral*, UTM, Springer-Verlag New York, Inc., 2000.
- [Cha] Chae, Soo B.: *Lebesgue Integration*, Springer-Verlag; 2nd Edition, 1995.
- [Con] Conway, J. B.: *A Course in Functional Analysis*, GTM, Springer-Verlag New York, Inc., 1985.
- [CuP] Curtain, R. F., Pritchard, A. J.: *Functional Analysis in Modern Applied Mathematics, Mathematics in Science and Engineering* 132, Academic Press Inc. (London) Ltd., 1977.
- [DaSe] Dangelo, F., Seyfried, M.: *Introductory Real Analysis*, Houghton Mifflin Company, 2000.

- [DeS] DePree J., Swartz, C.: *Introduction to Real Analysis*, New York, Toronto, Wiley, 1988.
- [deS] de Souza, P. N., Silva, J. N.: *Berkeley Problems in Mathematics*, 2nd edition, Springer, 2001.
- [EiMT] Eidelman, Y., Milman, V., Tzolomitis, A.: *Functional Analysis: An Introduction*, GSM 66, American Mathematical Society, Providence, Rhode Island, 2004.
- [EvG] Evans, L. C.: *Measure Theory and Fine Properties of Functions*, CRC Press, 1992.
- [Fla] Flatto, L.: *Advanced Calculus*, The Williams & Wilkins Company, 1976.
- [Fol] Folland, G. B.: *Real Analysis, Modern Techniques and Their Applications*, Pure & Applied Mathematics, A Wiley-Interscience Series of Texts, Monographs, and Tracts, John Wiley & Sons, Inc., 1984.
- [Frid] Fridy, J. A.: *Introductory Analysis: The Theory of Calculus* (2nd Edition), San Diego, Toronto, Harcourt Brace Jovanovich, Academic Press, 2000.
- [Frie] Friedman A.: *Foundations of Modern Analysis*, Holt, Rinehart and Winston, Inc., 1970.
- [GaN] Gaskill, H. S, Narayanaswami, P. P.: *Elements of Real Analysis*, Upper Saddle River, NJ, Prentice Hall, 1998.
- [Geo] George, C.: *Exercises in Integration*, Springer-Verlag New York, Inc., 1984.
- [Gro] Groetsch, C. W.: *Elements of Applicable Functional Analysis*, Marcel Dekker, Inc., New York and Basel, 1980.
- [Hil] Hille, E.: *Methods in Classical and Functional Analysis*, Addison-Wesley Publishing Company, Inc., 1972.
- [KolF] Kolmogorov, A. N., Fomin, S. V.: *Elements of the Theory of Functions and Functional Analysis*, Vol. 1 & 2, Graylock Press, 1957.

- [KuS] Kurtz, D. S., Swartz, C. W.: Theories of Integrations: The Integrals of Riemann, Lebesgue, Henstock-Kurzweil, and Mcshane, *Series in Real Analysis* 9, World Scientific Publishing Co. Pte. Ltd., 2004.
- [Nie] Nielsen, O. A.: An Introduction to Integration and Measure Theory, *Canadian Mathematical Society Series of Monographs and Advanced Texts*, John Wiley & Sons, Inc. 1997.
- [Pac] Packel, E. W.: *Functional Analysis: A Short Course*, Intext, Inc., 1974.
- [Pit] Pitt, H. R.: *Measure and Integration for Use*, Clarendon Press, Oxford, 1985.
- [Ros] Rosenlicht, M.: *Introduction to Analysis*, Scott, Foresman and Company, 1968
- [Roy] Royden, H. L.: *Real Analysis*, Prentice-Hall, Inc., 1998.
- [Rud] Rudin, W.: *Functional Analysis*, Second Edition, McGraw-Hill, Inc., 1991.
- [RyY] Rynne, B. P., Youngson, M. A.: *Linear Functional Analysis*, SUMS, Springer-Verlag London Limited, 2000.
- [Sax] Saxe, K., *Beginning Functional Analysis*, UTM, Springer-Verlag New York, Inc., 2002.
- [Sch] Schechter, M.: *Principles of Functional Analysis*, Academic Press, Inc., 1971.
- [Tor] Torchinsky, A.: *Real Variables*, Addison-Wesley Publishing Company, 1988.
- [Ves] Vestrup, E. W.: *The Theory of Measures and Integration*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., 2003.
- [War] Ward, T. B.: *Functional Analysis Lecture Notes*, https://archive.org/details/TB_Ward__Functional_analysis_lecture_notes/mode/2up.

- [You] Young, N.: *An Introduction to Hilbert Space*, Cambridge University Press, 1998.
- [Zim] Zimmer, R. J.: *Essential Results of Functional Analysis*, The University of Chicago Press, 1990.

About the Author

Jie Xiao

University Research Professor
Memorial University

- Education-Employment-Honor:
 - PhD Peking, 1992; Alexander von Humboldt Research Fellow, 1999-2000;
 - Memorial's Dean of Science Distinguished Scholar Medal, 2009;
 - Memorial's University Research Professor, 2016-present; Recognized on Stanford University List of World's Top 2% Scientists, 2020
- 199 Publications including:
 - 9 papers in *Adv. Math.*;
 - 4 papers in *Math. Ann.*;
 - 5 papers in *Calc Vari.*;
 - 2 papers in *J. Math. Pures Appl.*;
 - 1 paper in *J. Euro. Math. Soc.*;
 - 1 paper in *Ann. Inst. H. Poincar'e Anal. Non Lin'aire*;
 - 3 monographs;
 - 2 textbooks
- Services including:
 - Editor-in-Chief, *Advances in Analysis and Geometry*, 2018-present;
 - Editor-in-Chief of the *Canadian Mathematical Bulletin*, 2014-2019;
 - Associate Editor of the *Journal of Mathematical Analysis and Applications*, 2007-present

Index

- $\bar{\cdot}$ -measurable, 127
- \mathfrak{A} -ring, 110
- g -null set, 71
- m_g -measurable, 107
- m_g -null function, 98
- m_g -null set, 98
- $g_1 \otimes g_2$ -measure, 113
- absolutely convergent, 80, 256
- adjoint operator, 290, 311
- admissible sequences, 90
- affine subset, 227
- Alaoglu's theorem, 286
- algebraic closure, 33, 63
- algebraically isomorphic, 233
- angle, 301
- antisymmetry, 8
- Arzela-Ascoli theorem, 210
- Ascoli's theorem, 212
- associativity, 10
- axiom of choice, 10
- Banach algebra, 259, 292
- Banach fixed point theorem, 212
- Banach-Zarecki theorem, 151
- basis, 228, 308
- Bessel's inequality, 308
- Bonnet's form, 77
- braces, 1
- cardinality, 6
- Cartesian product, 3, 238
- Cauchy's convergence criterion, 45
- Cauchy's criterion, 29
- Cauchy's sequence, 17
- Cauchy-Schwarz inequality, 34
- chain, 8
- Chebychev polynomials, 306
- closed graph theorem, 266
- closed set, 13
- cluster point, 13
- codimension, 230
- commutativity, 10
- compact metric space, 188, 204
- compact self-adjoint, 325
- comparison test, 46
- completeness of \mathbf{R} , 18
- complex linear space, 227
- composition, 69
- conditionally convergent, 80
- connected metric space, 207
- continuous spectrum, 322
- convergence theorem in $LRS_g^p(\mathbf{R})$, 122
- convolution, 50, 292
- countable additivity, 89
- countably additive, 143
- countably subadditive, 110

- Darboux characterization of integrability, 32
Darboux's criterion, 29
DeMorgan law, 2
differential operator, 322
Dirac sequence, 50
direct sum, 302
disconnected, 179
distributivity, 10
domain, 4
dominated convergence, 100
- eigenspace, 326
eigenvalue, 322
eigenvector, 322
equivalence relation, 3
equivalent, 218, 234
existence of inverses, 10
existence of neutrals, 10
- factor space, 230
Fatou's lemma, 104
finite intersection property, 285
finite rank, 321
finite subcover, 187
first and second mean value theorem, 75
first category, 194
four one-sided derivatives, 144
Fourier series, 309
Fourier transform, 318
Fredholm integral equation of the second kind, 213
Fubini-Tonelli's theorem, 114
fundamental theorem of Lebesgue integration, 151
- Gamma function, 49
- general fixed point theorem, 214
general form of Minkowski's inequality, 120
generalized Cauchy-Schwartz inequality, 315
generalized mean value theorem, 41
Gram-Schmidt process, 305
- Hölder inequality, 118
Hahn-Banach extension theorem, 244
Hanner's inequality, 273
Hardy-Littlewood's maximal theorem, 166
Hermite polynomials, 306
homogeneity, 228
- image, 4
imaginary, 2
infima, 9
infinite-dimensional, 228
infinitely differentiable, 226
injective, 4
inner product, 297
integrand, 61
integrator, 61
interior point, 13
inverse mappings, 266
isometric, 234
isometrically isomorphic, 218
- Jacobi polynomials, 306
- kernel, 51
Krein's extension theorem, 253
- Laguerre polynomials, 306
Laplace transform, 55
layer cake representation, 125

- Lebesgue's decomposition theorem, 170
- Lebesgue's differentiation theorem, 165
- Lebesgue's integration-by-part formula, 157
- Lebesgue's property, 193
- Lebesgue's theorem on the derivative, 146
- Lebesgue-Radon-Stieltjes space, 121
- Legendre polynomials, 306
- linear span, 227
- Lipschitz class, 148
- Lipschitz condition, 216
- LUB-axiom, 11
- maximal ideal, 10
- Minkowski's inequality, 118
- monotone convergence, 103
- monotone convergence theorem, 17
- nested downward (respectively upward) sequence, 13
- nonnegative operator, 315
- normal operator, 319
- nowhere dense, 194
- objects, 1
- open mapping theorem, 264
- open set, 13
- orthogonal, 301
- orthogonal complement, 301
- orthogonal projection, 321
- orthonormal basis, 308
- parallelogram law, 298
- Parseval's identity, 309
- partition, 21
- Picard's existence and uniqueness theorem, 216
- point spectrum, 322
- polynomial approximation, 50
- pre-image, 4
- product topological space, 285
- projection, 320
- projection operator, 320
- range, 4
- real linear space, 227
- rectangle, 113
- rectifiable, 83
- refinement, 30
- reflexive, 243, 287
- reflexivity, 3, 8
- relation, 3
- residual spectrum, 322
- Riemann integrable, 22
- Riemann sum, 22
- Riemann-Stieltjes integrable, 61
- Riemann-Stieltjes sum, 61
- Riesz representation theorem, 277, 303
- second category, 194
- self-adjoint or symmetric operator, 313
- self-reflexive, 304
- semi-norm, 228
- similar, 218
- simple sets, 88
- spectral radius, 322
- spectrum, 322
- strongly convergent, 283
- summable functions, 90
- suprema, 9

- surjective, 4
- symmetry, 3
- The Bolzano-Weierstrass property,
14, 184
- The bounded sequence property,
184
- The Heine-Borel property, 14, 184
- The nested interval property, 14
- The nested set property, 184
- topologically isomorphic, 234
- topologically isomorphic or home-
omorphic, 218
- total/linear order, 8
- totally bounded, 189
- totalness, 8
- transitivity, 3, 8
- triangle inequality, 228
- Tychonoff's theorem, 285
- uniform boundedness principle,
261
- uniform convergence, 208
- uniform limit, 289
- uniformly continuous, 36
- uniformly isomorphic, 218
- unit sphere, 237
- unitary, 317
- unitary operator, 318
- variation, 57
- Volterra integral equation, 215
- weak compactness, 287
- weak topology, 282
- weak* closed, 284
- weak* compact, 284
- weak* convergent, 283
- weak* open, 284
- weakly closed, 282
- weakly compact, 282
- weakly convergent, 283
- weakly open, 282
- Weierstrass approximation, 52
- Zorn's lemma, 9

Integral and Functional Analysis

(Updated Edition)



Jie Xiao



nova
science publishers
www.novapublishers.com

ISBN 978-1-53619-280-3



9 781536 192803